Quantum Simulation of Enzyme Catalysis

Thesis by
Nicholas Boekelheide

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2013

(Defended June 6, 2013)

Acknowledgements

First, thank you to Tom Miller, my advisor. His support, guidance, and encouragement have contributed a great deal to my development as a scientist over the last five years. He taught me how to think about the problems I find most interesting.

I would also like to thank the chemistry division faculty. Particularly, I thank my thesis committee, Harry Gray, Rudy Marcus, and Shu-ou Shan for their time, advice and help planning for the future. I also thank Mitchio Okumura and Doug Rees. I appreciated working with them my first year and their interest in what I've done since. Also, I remember Aron Kuppermann for his articulate, detailed teaching of quantum mechanics and friendly conversations outside of class.

Thank you to the division administrative staff, particularly Priscilla Boon who does so much to support the Miller group and the chemical physics subdivision.

I appreciate the friendship and support of the Miller Group. Bin, Artur, Jason, Josh, Taylor, Connie, Mike, Michiel, Nandini, Romeila, Kuba, Fran, and Frank, I couldn't have had better officemates. Feedback on talks and papers and help developing my ideas taught me so much. I've enjoyed all the scientific discussions, walks for coffee, and fun we've had. Thank you as well to my friends outside the Miller Group who were there to help me unwind.

I also thank the teachers and students who I mentored with the Caltech Classroom Connection, Christine Kye, Phil Sang, Darius Hines, and Kyle Priver. Additionally, I thank James Maloney and Sherry Tsai, the coordinators of the Caltech Classroom Connection. I enjoyed our work together and the opportunity to explore education in the sciences beyond Caltech.

Thank you to my family—my parents, my brothers and the Bergmanns. I'm not sure what you guys thought of what Kristin and I did, but we knew you were there for us, and that helped.

And to Kristin, for doing this together.

Quantum Simulation of Enzyme Catalysis

by

Nicholas Boekelheide

In Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Abstract

Separating the dynamics of variables that evolve on different timescales is a common assumption in exploring complex systems, and a great deal of progress has been made in understanding chemical systems by treating independently the fast processes of an activated chemical species from the slower processes that proceed activation. Protein motion underlies all biocatalytic reactions, and understanding the nature of this motion is central to understanding how enzymes catalyze reactions with such specificity and such rate enhancement. This understanding is challenged by evidence of breakdowns in the separability of timescales of dynamics in the active site form motions of the solvating protein. Quantum simulation methods that bridge these timescales by simultaneously evolving quantum and classical degrees of freedom provide an important method on which to explore this breakdown. In the following dissertation, three problems of enzyme catalysis are explored through quantum simulation.

In Chapter 1, we address the role of protein motion in enzyme catalysis. By quantifying and distinguishing between statistical and dynamical correlations in catalysis by dihydrofolate reductase, we are able to define hierarchies of motion in enzyme catalysis and confirm that nanometer-scale protein fluctuations statistically gate the barrier for the intrinsic reaction are not dynamically correlated during the intrinsic reaction event. The manuscript associated with this study is published in *PNAS*, 108, 16159 (2011).

In Chapter 2, we explore the mechanistic origin of temperature independent kinetic isotope effects in enzyme catalysis. By enabling the characterization of the ensemble of reactive paths, quantum simulations reveal transient compression of the donor and acceptor atoms that is independent of the mass of the transferred hydride, thus strengthening the mechanistic interpretation of this general feature of enzyme catalysis.

Chapter 3 explores the nature of protein motion in lactate dehydrogenase. The mechanism and energetics captured in simulations is shown to depend on treatment of the potential energy in the active site. A method is established to explore the nature of coupling between protein motion and the dynamics of the active site in the context of previous results.

Contents

List of Figures

List of Tables

Chapter 1

Dynamics and Dissipation in Enzyme Catalysis

We use quantized molecular dynamics simulations to characterize the role of enzyme vibrations in facilitating dihydrofolate reductase hydride transfer. By sampling the full ensemble of reactive trajectories, we are able to quantify and distinguish between statistical and dynamical correlations in the enzyme motion. We demonstrate the existence of non-equilibrium dynamical coupling between protein residues and the hydride tunneling reaction, and we characterize the spatial and temporal extent of these dynamical effects. Unlike statistical correlations, which give rise to nanometer-scale coupling between distal protein residues and the intrinsic reaction, dynamical correlations vanish at distances beyond 4-6 Å from the transferring hydride. This work suggests a minimal role for non-local vibrational dynamics in enzyme catalysis, and it supports a model in which nanometer-scale protein fluctuations statistically modulate—or gate—the barrier for the intrinsic reaction.

1.1 Introduction

Protein motions are central to enzyme catalysis, with conformational changes on the micro- and milli-second timescale well-established to govern progress along the catalytic cycle [1, 2]. Less is known about the role of faster, atomic-scale fluctua-

tions that occur in the protein environment of the active site. The textbook view of enzyme-catalyzed reaction mechanisms neglects the functional role of such fluctuations and describes a static protein environment that both scaffolds the active site region and reduces the reaction barrier [3]. This view has grown controversial amid evidence that active site chemistry is coupled to motions in the enzyme [4, 5, 6], and it has been explicitly challenged by recent proposals that enzyme-catalyzed reactions are driven by vibrational excitations that channel energy into the intrinsic reaction coordinate [7, 8] or promote reactive tunneling [9, 10]. In the following, we combine quantized molecular dynamics and rare-event sampling methods to reveal the mechanism by which protein motions couple to reactive tunneling in dihydrofolate reductase and to clarify the role of non-equilibrium vibrational dynamics in enzyme catalysis.

Manifestations of enzyme motion include both *statistical* and *dynamical* correlations. Statistical correlations are properties of the equilibrium ensemble and describe, for example, the degree to which fluctuations in the spatial position of one atom are influenced by fluctuations in another; these correlations govern the free energy (FE) landscape and determine the transition state theory kinetics of the system [6]. Dynamical correlations are properties of the time-evolution of the system and describe coupling between inertial atomic motions, as in a collective vibrational mode. Compelling evidence for long-ranged (i.e., nanometer-scale) networks of statistical correlations in enzymes emerges from genomic analysis [11], molecular dynamics simulations [11, 12, 13], and kinetic studies of double-mutant enzymes [14, 15]. But the role of dynamical correlations in enzyme catalysis remains unresolved [16, 4, 5, 7], with experimental and theoretical results suggesting that the intrinsic reaction is activated by vibrational modes involving the enzyme active site [9, 17, 18] and more distant protein residues [19, 7, 8]. The degree to which enzyme-catalyzed reactions are coupled to the surrounding protein environment, and the lengthscales and timescales over which such couplings persist, are central questions in the understanding, regulation, and *de novo* design of biological catalysts [20].

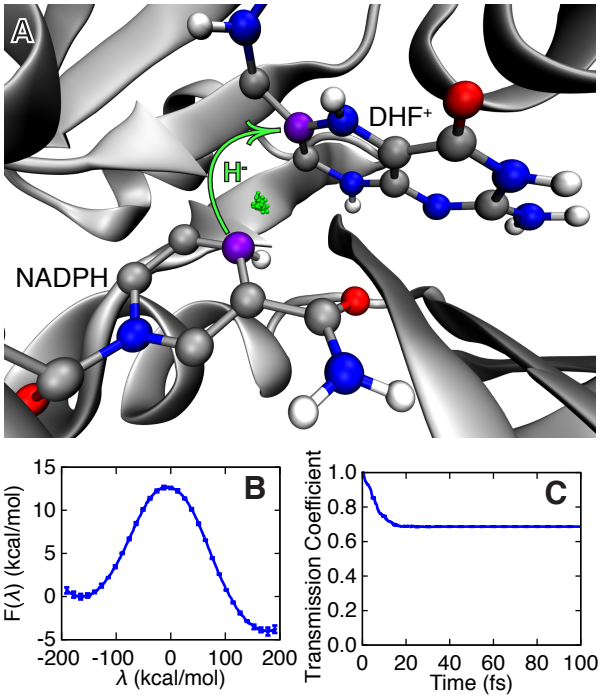

Figure 1.1: The hydride transfer reaction catalyzed by DHFR. (*A*) The active site with the hydride (green) shown in the ring-polymer representation of the quantized MD and the donor and acceptor C atoms in purple. (*B*) The quantized free energy profile for the reaction. (*C*) The time-dependent transmission coefficient corresponding to the dividing surface at $\lambda(\mathbf{x}) = -4.8$ kcal/mol.

Escherichia coli dihydrofolate reductase (DHFR) is an extensively studied prototype for protein motions in enzyme catalysis. It catalyzes reduction of the 7,8-dihydrofolate (DHF) substrate via hydride transfer from the nicotinamide adenine dinucleotide phosphate (NADPH) cofactor (Figure 1.1A). We investigate this intrinsic reaction using ring polymer molecular dynamics (RPMD) [21, 22], a recently developed path-integral method that enables inclusion of nuclear quantization effects, such as the zero-point energy and tunneling, in the dynamics of the transferring hydride. RPMD simulations with over 14,000 atoms are performed using explicit solvent and using an empirical valence bond (EVB) potential to describe the potential energy surface for the transferring hydride; the EVB potential

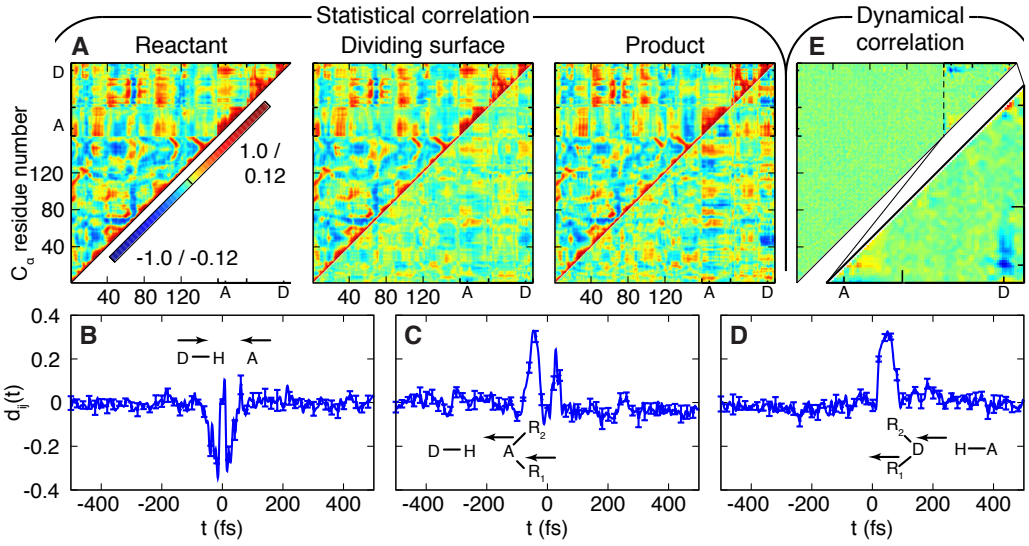

Figure 1.2: Statistical and dynamical correlations among enzyme motions during the intrinsic reaction. (*A*) (Upper triangles) The covariance c_{ij} among position fluctuations in DHFR, plotted for the reactant, dividing surface, and product regions. Protein residues are indexed according to PDB 1RX2; substrate and cofactor regions are indicated by the hydride acceptor A and donor D atoms, respectively. (Lower triangles) The difference with respect to the plot for the reactant basin. (*B-D*) The dynamical correlation measure $d_{ij}(t)$ for (*B*) the donor and acceptor atom pair, (*C*) the substrate-based C_7 and acceptor atom pair, and (*D*) the cofactor-based C_{N3} and donor atom pair. Results for additional atom pairs are presented in Figure A.8. (*E*) (Upper triangle) The integrated dynamical correlation measure d_{ij}, indexed as in (*A*). Significant dynamical correlations appear primarily in the substrate and cofactor regions, which are enlarged in the lower triangle.

is obtained from an effective Hamiltonian matrix, with diagonal elements ($V_\mathrm{r}(\mathbf{x})$ and $V_\mathrm{p}(\mathbf{x})$) corresponding to the potential energy for the reactant and product bonding connectivities and with the constant off-diagonal matrix element fit to the experimental rate [23, 11]. The vector $\mathbf{x}$ includes the position of the quantized hydride and all classical nuclei in the system. The thermal reaction rate is calculated from the product of the Boltzmann-weighted activation FE and the reaction transmission coefficient [22], both of which are calculated in terms of the dividing surface $\lambda(\mathbf{x}) = -4.8$ kcal/mol where $\lambda(\mathbf{x}) = V_\mathrm{r}(\mathbf{x}) - V_\mathrm{p}(\mathbf{x})$. The FE sur-

face $F(\lambda)$ is obtained using over 120 ns of RPMD sampling (Figure 1.1B), and the transmission coefficient is obtained from over 5000 RPMD trajectories that are released from the Boltzmann distribution constrained to the dividing surface (Figure 1.1C). In contrast to mixed quantum-classical and transition state theory methods, RPMD yields reaction rates and mechanisms that are formally independent of the choice of dividing surface or any other reaction coordinate assumption [22]. Furthermore, the RPMD method enables generation of the ensemble of reactive, quantized molecular dynamics trajectories, which is essential for the following analysis of dynamical correlations. Calculation details, including a description of the rare-event sampling methodology used to generate the unbiased ensemble of reactive trajectories [24, 25], are provided in below.

1.2 Results and Discussion

The time-dependence of the transmission coefficient in Figure 1.1C confirms that reactive trajectories commit to the reactant or product basins within 25 fs. The near-unity value of this transmission coefficient at long times indicates that recrossing of the dividing surface in reactive trajectories is a modest effect, although it is fully accounted for in this study, and it confirms that the collective variable $\lambda(\mathbf{x})$ provides a good measure of progress along the intrinsic reaction. We find that quantization of the hydride lowers the FE barrier by approximately 3.5 kcal/mol (Figure A.5), in agreement with earlier work [26, 27].

Statistical correlations among the protein and enzyme active site coordinates are shown in Figure 1.2A. The normalized covariance among atom position fluctuations, $c_{ij} = C_{ij}/(C_{ii}C_{jj})^{1/2}$ such that

$$C_{ij} = \langle (\mathbf{x}_i - \langle \mathbf{x}_i \rangle) \cdot (\mathbf{x}_j - \langle \mathbf{x}_j \rangle) \rangle \,, \tag{1.1}$$

is plotted for the Boltzmann distribution in the reactant, dividing surface, and product regions. The figure shows correlations among the protein α-carbons and the heavy atoms of the substrate and cofactor; the corresponding all-atom corre-

lation plots are provided in Figure A.6. As has been previously emphasized [11], structural fluctuations in the active site and distal protein residues are richly correlated within each region, which contributes to non-additive effects in the kinetics of DHFR mutants [28, 14]. Furthermore, the network of correlations varies among the three ensembles, indicating that fluctuations in distal protein residues respond to the adiabatic progress of the hydride from reactant to product. However, these time-averaged quantities do not address the role of dynamical correlations between the transferring hydride and its environment, which depend on the hierarchy of timescales for motion in the system.

To characterize dynamical correlations in the intrinsic reaction, we introduce a measure of velocity cross-correlations in the reactive trajectories, $d_{ij}(t) = D_{ij}(t)/(D_{ii}(t)D_{jj}(t))^{1/2}$ such that

$$D_{ij}\left(t\right) = \left\langle \mathbf{v}_i \cdot \mathbf{v}_j \right\rangle_t . \tag{1.2}$$

Here, $\langle ... \rangle_t$ denotes an average over the non-equilibrium ensemble of phase-space points that lie on reactive trajectories which crossed the dividing surface some time t earlier and subsequently terminate in the product basin. This quantity, which vanishes for the equilibrium ensemble, reports on the degree to which atoms move in concert during the intrinsic reaction step. Figure 1.2B-D show $d_{ij}(t)$ for several atomic pairs in the active site. Negative dynamical correlations are seen between the donor and acceptor C atoms (Figure 1.2B), which move in opposite directions (first approaching each other, then moving apart) during the hydride transfer. Similarly, positive correlations are seen between atom pairs on the cofactor (Figure 1.2C) and on the substrate (Figure 1.2D) which move in concert as the hydride is transferred. In each case, the primary features of the correlation decay within $\tau = 100$ fs.

Figure 1.2E summarizes the extent of dynamical correlations throughout the enzyme system in terms of $d_{ij} = \int_{-\tau}^{\tau} d_{ij}(t)dt$. Only atoms in the substrate and cofactor regions (Figure 1.2E, lower triangle) and a small number of protein atoms in

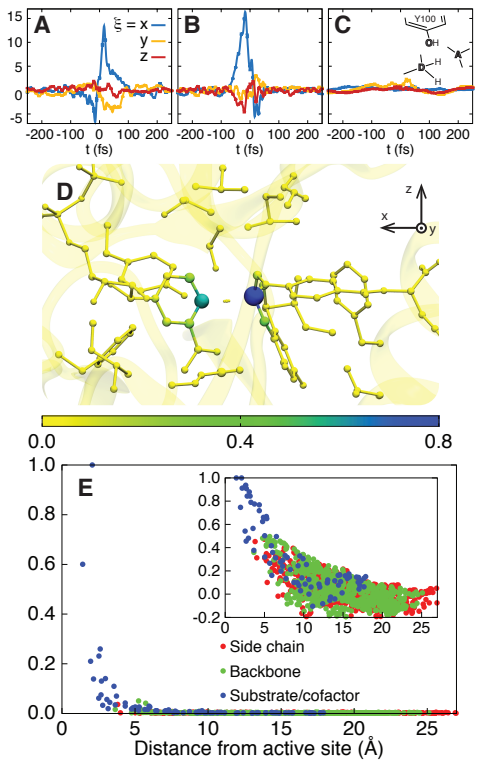

Figure 1.3: The dynamical correlation measure $f_i^{\xi}(t)$, plotted for (A) the donor atom, (B) the acceptor atom, and (C) the side-chain O atom in the Y100 residue of the active site. (D) The size and color of atoms in the active site region are scaled according to the integrated dynamical correlation measure, f_i. (E) (main panel) The integrated dynamical correlation measure, f_i, as a function of the distance of atom i from the midpoint of the donor and acceptor atoms. (inset) The statistical correlation measure, $\bar{c}_i$, is similarly presented. Atoms corresponding to the protein side chains, the protein backbone, and the substrate/cofactor regions are indicated by color. Values presented in part A are in units of nm/ps, and values in parts D and E are normalized to a maximum of unity. The estimated error in part E is smaller than the dot size.

the active site region exhibit appreciable signal. The same conclusions are reached upon integrating the absolute value of the $d_{ij}(t)$ (Figure A.7), emphasizing that this lack of signal in the protein residues is not simply due to the time integration. Instead, Figure 1.2E reveals that the dynamical correlations between distal protein residues and the intrinsic reaction *do not exist on any timescale*. We also provide measures for dynamical correlations among perpendicular motions (Figure A.9)

and for dynamical correlations that are non-local in time (Figure A.8), but the following conclusion is unchanged. The extensive network of statistical correlations (Figure 1.2A) is neither indicative of, nor accompanied by, an extensive network of dynamical correlations during the intrinsic reaction (Figure 1.2E).

A combined measure of the dynamical correlation between a given atom and the intrinsic reaction event can be obtained from the non-equilibrium ensemble average of velocities in the reactive trajectories. Specifically, we consider $f_i^{\xi}(t) = \left\langle v_i^{\xi} \cdot \Delta\lambda(\mathbf{x}) \right\rangle_t$, where $\xi \in \{\mathrm{x}, \mathrm{y}, \mathrm{z}\}$ indicates the component of the velocity, the filter $\Delta\lambda(\mathbf{x}) = \left(\bar{\lambda} - |\lambda(\mathbf{x})|\right)/\bar{\lambda}$ selects configurations in the region of the dividing surface, and $\bar{\lambda} = 177$ kcal/mol is the average magnitude of $\lambda(\mathbf{x})$ in the reactant and product regions. Each component of $\mathbf{f}_i(t)$ vanishes trivially at equilibrium. Figure 1.3A-C presents the measure for various atoms in the active site region. The donor and acceptor C atoms (Figure 1.3A and B) are both strongly correlated with the dynamics of the intrinsic reaction, whereas the O atom in the Y100 residue of the active site (Figure 1.3C) reveals smaller, but non-zero, signatures of dynamical correlation. Figure 1.3D presents $f_i = \int_{-\tau}^{\tau} |\mathbf{f}_i(t)|^2 dt$ for each atom, summarizing the degree to which all atoms in the active site exhibit dynamical correlations, and Figure 1.3E compares the correlation lengthscales in the enzyme. The main panel in Figure 1.3E presents f_i as a function of the distance of heavy atoms from the midpoint of the hydride donor and acceptor, and the inset similarly presents the distance dependence of the statistical correlation measure $\bar{c}_i = (c_{i\mu} + c_{i\nu})/2$, where c_{ij} is defined previously and where indices μ and ν label the donor and acceptor carbon atoms, respectively. Whereas the statistical correlations reach the nanometer lengthscale and involve the protein environment, dynamical correlations are extremely local in nature and primarily confined to the enzyme substrate and cofactor.

Figure 1.4 illustrates that dynamical correlations in the intrinsic reaction are limited by disparities in the relative timescales for enzyme motion. The figure presents two-dimensional projections of the FE surface, $F(\lambda, \Theta_\alpha)$, where $\alpha \in \{1, 2\}$, $\Theta_1(\mathbf{x})$ is the distance between hydride donor and acceptor atoms, and $\Theta_2(\mathbf{x})$ is the

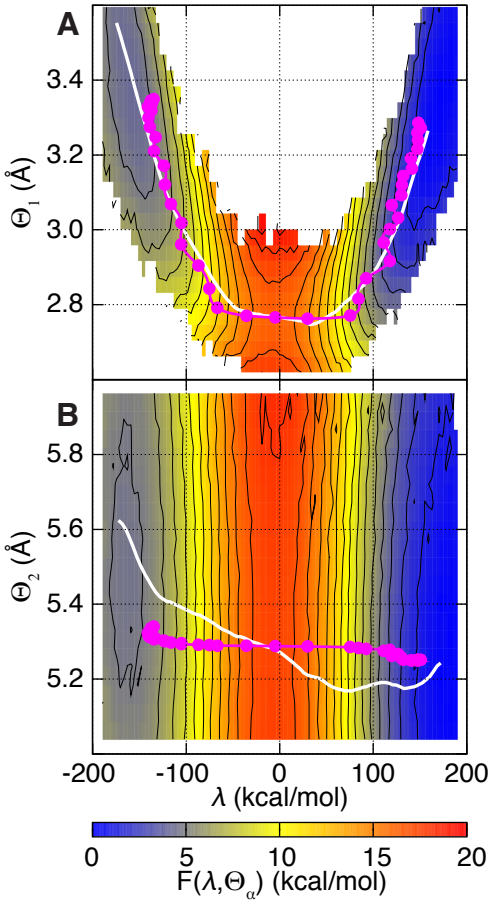

Figure 1.4: Minimum free energy pathways (s, white) and the mean pathway of the reactive trajectories (σ, magenta) overlay two-dimensional projections of the free energy landscape, $F(\lambda, \Theta_\alpha)$. ($A$) $F(\lambda, \Theta_1)$, where Θ_1 is the distance between the hydride donor and acceptor atoms. (B) $F(\lambda, \Theta_2)$, where Θ_2 is the distance between side-chain atoms I14 C_δ and Y100 O in the active site residues. The non-zero slope in s and σ indicates statistical and dynamical correlations, respectively.

separation between active site protein atoms I14 C_δ and Y100 O (side chain). Overlaid on the surfaces are the minimum FE pathway between the reactant and product basins, s, and the time-parameterized pathway followed by the ensemble of reactive trajectories, $\sigma = (\langle\lambda(\mathbf{x})\rangle_t, \langle\Theta_\alpha(\mathbf{x})\rangle_t)$. Non-zero slope in s indicates statistical correlation of Θ_α with λ, whereas the same feature in σ indicates that the dynamics of Θ_α and λ are dynamically correlated. Figure 1.4A confirms that the

donor-acceptor distance is both statistically and dynamically correlated with the intrinsic reaction. In contrast, Figure 1.4B reveals significant statistical correlation between Θ_2 and the intrinsic reaction, but the reactive trajectories traverse the dividing surface region on a timescale that is too fast to dynamically couple to the protein coordinate.

1.3 Concluding Remarks

The results presented here complement previous theoretical efforts to illuminate the role of protein motions in enzyme catalysis. For example, Neria and Karplus [29] used transmission coefficient calculations and constrained MD trajectories to determine that the protein environment in triosephosphate isomerase (TIM) is essentially rigid (i.e., dynamically unresponsive) on the timescale of the intrinsic reaction dynamics; this finding is consistent with the lack of long-lengthscale dynamical correlations found in the current study. Furthermore, Pu et al. [30] and Cui and Karplus [31] both demonstrated that quasi-classical tunneling coefficients for hydrogen transfer evaluated at instantaneous enzyme configurations in the transition state region fluctuate significantly with donor-acceptor motions and other local active-site vibrations, which is likely consistent with the direct observation of short-lengthscale dynamical correlations reported here. However, by using quantized MD to sample the ensemble of reactive trajectories in DHFR catalysis and to perform non-equilibrium ensemble averages that directly probe dynamical correlation, we provide a framework for strengthening and generalizing these earlier analyses. In particular, the current approach avoids transition state theory approximations by providing a rigorous statistical mechanical treatment of the ensemble of reactive trajectories, it allows for the natural characterization of lengthscales and timescales over which dynamical correlations persist, and it seamlessly incorporates dynamical effects due to both nuclear quantization and trajectory recrossing. We expect that the techniques developed here to prove useful in future studies of dynamics in other enzymes, which will be necessary to

confirm the generality of the conclusions drawn here.

1.4 Methods

1.4.1 Calculation details

All simulations are performed using a modified version of the Gromacs-4.0.7 molecular dynamics package [32]. Further calculation details regarding the potential energy surface, the system initialization and equilibration protocol, free-energy sampling, and dividing surface sampling are provided in Appendix A.

1.4.2 Ring polymer molecular dynamics

The RPMD equations of motion [21] used to simulate the dynamics of DHFR are

$$\ddot{\mathbf{q}}_j = \frac{1}{nm_n}\left[k_n(\mathbf{q}_{j+1}+\mathbf{q}_{j-1}-2\mathbf{q}_j) - \nabla_{\mathbf{q}_j}U(\mathbf{q}_j,\mathbf{Q}_1,\ldots,\mathbf{Q}_N)\right], \qquad j=1,...,n \tag{1.3}$$

$$\ddot{\mathbf{Q}}_k = -\frac{1}{nM_k}\sum_{j=1}^{n}\nabla_{\mathbf{Q}_k}U(\mathbf{q}_j,\mathbf{Q}_1,\ldots,\mathbf{Q}_N), \qquad k=1,...,N \tag{1.4}$$

where $U(\mathbf{q},\mathbf{Q}_1,\ldots,\mathbf{Q}_N)$ is the potential energy function for the system, $n = 32$ is the number of ring polymer beads used to quantize the hydride, $\mathbf{q}_j$ and m_n are the position and mass of the j^{th} ring polymer bead, and $\mathbf{q}_0 = \mathbf{q}_n$. Similarly, N is the number of classical nuclei in the system, and $\mathbf{Q}_k$ and M_k are the position and mass of the k^{th} classical atom, respectively. The inter-bead force constant is $k_n = m_{\mathrm{H}}n^2/(\beta\hbar)^2$, where $m_{\mathrm{H}} = 1.008$ amu is the mass of the hydride and $\beta = (k_{\mathrm{B}}T)^{-1}$ is the reciprocal temperature; a temperature of $T = 300$ K is used throughout the study. For dynamical trajectories, RPMD prescribes that $m_n = m_{\mathrm{H}}/n$.

1.4.3 Calculating the statistical correlation functions, c_{ij}

In Figure 1.3A, equilibrium ensemble averages are presented for the system in the reactant region, the dividing surface region, and the product region. These

ensemble averages are strictly defined using

$$\begin{aligned}\langle A\rangle_{\lambda^*} &= \int_{\lambda^*-\delta\lambda}^{\lambda^*+\delta\lambda} d\lambda' P(\lambda') \int d\mathbf{q}_1 \ldots \int d\mathbf{q}_n \int d\mathbf{Q}_1 \ldots \int d\mathbf{Q}_N \\ &\times \delta(\lambda' - \lambda(\mathbf{x})) A(\mathbf{q_1}, \ldots, \mathbf{q_n}, \mathbf{Q}_1, \ldots, \mathbf{Q}_N), \end{aligned} \tag{1.5}$$

where $P(\lambda) = \exp(-\beta F(\lambda))/\int d\lambda' \exp(-\beta F(\lambda'))$, and $F(\lambda)$ is calculated using umbrella sampling, as described in the SI. For the ensembles in the reactant, dividing surface, and product regions, we employ λ^* = -181 kcal/mol, -7 kcal/mol, and 169 kcal/mol, respectively, and $\delta\lambda$ = 2.5 kcal/mol.

1.4.4 The transition path ensemble

Reactive trajectories are generated through forward- and backward-integration of initial configurations drawn from the dividing surface ensemble with initial velocities drawn from the Maxwell-Boltzmann distribution. Reactive trajectories correspond to those for which forward- and backward-integrated half-trajectories terminate on opposite sides of the dividing surface. From the 10,500 half-trajectories that are initialized on the dividing surface (i.e., 5,250 possible reactive trajectories), over 3,000 reactive RPMD trajectories are obtained. For analysis purposes, the integration of these reactive trajectories was continued for a total length of one ps in both the forward and backward trajectories.

The reactive trajectories that are initialized from the equilibrium Boltzmann distribution on the dividing surface must be reweighted to obtain the unbiased ensemble of reactive trajectories (i.e., the transition path ensemble) [25, 33, 24]. A weighting term is applied to each trajectory α, correctly accounting for the recrossing and for the fact that the trajectories are performed in the microcanonical ensemble [24],

$$w_\alpha = \left(\sum_{\text{intersections } i} \left| \dot{\lambda}_i \right|^{-1} \right)^{-1} \tag{1.6}$$

where the sum includes all instances in which trajectory α crosses the dividing surface, and $\dot{\lambda}_i$ is the velocity in the collective variable at crossing event i. We

find that the relative statistical weight of all reactive trajectories that recross the dividing surface is 1.6%, emphasizing that recrossing does not play a large role in the current study.

References

[1] K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hubner, and D. Kern. Intrinsic motions along an enzymatic reaction trajectory. *Nature*, 450:838–844, 2007.

[2] D. D. Boehr, D. McElheny, H. J. Dyson, and P. E. Wright. Millisecond timescale fluctuations in dihydrofolate reductase are exquisitely sensitive to the bound ligands. *Proc. Natl. Acad. Sci. U.S.A.*, 107(4):1373–1378, 2010.

[3] Paul F. Cook and W. W. Cleland. *Enzyme kinetics and mechanism*. Garland Science, New York, 2007.

[4] S. C. L. Kamerlin and A. Warshel. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins*, 78(6):1339–1375, 2010.

[5] V. C. Nashine, S. Hammes-Schiffer, and S. J. Benkovic. Coupled motions in enzyme catalysis. *Curr. Opin. Chem. Biol.*, 14(5):644–651, 2010.

[6] M. Garcia-Viloca, J. Gao, M. Karplus, and D. G. Truhlar. How enzymes work: Analysis by modern rate theory and computer simulations. *Science*, 303(5655):186–195, 2004.

[7] S. D. Schwartz and V. L. Schramm. Enzymatic transition states and dynamic motion in barrier crossing. *Nature Chem. Biol.*, 5(8):552–559, 2009.

[8] S. Q. Machleder, J. R. E. T. Pineda, and S. D. Schwartz. On the origin of the

chemical barrier and tunneling in enzymes. *J. Phys. Org. Chem.*, 23(7):690–695, 2010.

[9] L. Masgrau, A. Roujeinikova, L. O. Johannissen, P. Hothi, J. Basran, K. E. Ranaghan, A. J. Mulholland, M. J. Sutcliffe, N. S. Scrutton, and D. Leys. Atomic description of an enzyme reaction dominated by proton tunneling. *Science*, 312(5771):237–241, 2006.

[10] Z. D. Nagel and J. P. Klinman. A 21st century revisionist's view at a turning point in enzymology. *Nature Chem. Biol.*, 5(8):543–550, 2009.

[11] P. K. Agarwal, S. R. Billeter, P. T. R. Rajagopalan, S. J. Benkovic, and S. Hammes-Schiffer. Network of coupled promoting motions in enzyme catalysis. *Proc. Natl. Acad. Sci. U.S.A.*, 99(5):2794–2799, 2002.

[12] J. Y. Pang, J. Z. Pu, J. L. Gao, D. G. Truhlar, and R. K. Allemann. Hydride transfer reaction catalyzed by hyperthermophilic dihydrofolate reductase is dominated by quantum mechanical tunneling and is promoted by both inter- and intramonomeric correlated motions. *J. Am. Chem. Soc.*, 128(24):8015–8023, 2006.

[13] T. H. Rod, J. L. Radkiewicz, and C. L. Brooks. Correlated motion and the effect of distal mutations in dihydrofolate reductase. *Proc. Natl. Acad. Sci. U.S.A.*, 100(12):6980–6985, 2003.

[14] K. F. Wong, T. Selzer, S. J. Benkovic, and S. Hammes-Schiffer. Impact of distal mutations on the network of coupled motions correlated to hydride transfer in dihydrofolate reductase. *Proc. Natl. Acad. Sci. U.S.A.*, 102(19):6807–6812, 2005.

[15] L. Wang, S. Tharp, T. Selzer, S. J. Benkovic, and A. Kohen. Effects of a distal mutation on active site chemistry. *Biochemistry*, 45(5):1383–1392, 2006.

[16] S. J. Benkovic and S. Hammes-Schiffer. Biochemistry—Enzyme motions inside and out. *Science*, 312(5771):208–209, 2006.

[17] C. R. Pudney, S. Hay, C. Levy, J. Y. Pang, M. J. Sutcliffe, D. Leys, and N. S. Scrutton. Evidence to support the hypothesis that promoting vibrations enhance the rate of an enzyme catalyzed h-tunneling reaction. *J. Am. Chem. Soc.*, 131(47):17072–17073, 2009.

[18] E. J. Loveridge, L. H. Tey, and R. K. Allemann. Solvent effects on catalysis by *Escherichia coli* dihydrofolate reductase. *J. Am. Chem. Soc.*, 132(3):1137–1143, 2010.

[19] S. Saen-Oon, S. Quaytman-Machleder, V. L. Schramm, and S. D. Schwartz. Atomic detail of chemical transformation at the transition state of an enzymatic reaction. *Proc. Natl. Acad. Sci. U.S.A.*, 105(43):16543–16548, 2008.

[20] P. Ball. Enzymes—By chance, or by design? *Nature*, 431(7007):396–397, 2004.

[21] I. R. Craig and D. E. Manolopoulos. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *J. Chem. Phys.*, 121(8):3368–3373, 2004.

[22] I. R. Craig and D. E. Manolopoulos. A refined ring polymer molecular dynamics theory of chemical reaction rates. *J. Chem. Phys.*, 123(3):034102, 2005.

[23] A. Warshel and R. M. Weiss. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Am. Chem. Soc.*, 102(20):6218–6226, 1980.

[24] G. Hummer. From transition paths to transition states and rate coefficients. *J. Chem. Phys.*, 120(2):516–523, 2004.

[25] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.*, 53:291–318, 2002.

[26] P. K. Agarwal, S. R. Billeter, and S. Hammes-Schiffer. Nuclear quantum effects and enzyme dynamics in dihydrofolate reductase catalysis. *J. Phys. Chem. B*, 106(12):3283–3293, 2002.

[27] M. Garcia-Viloca, D. G. Truhlar, and J. L. Gao. Reaction-path energetics and kinetics of the hydride transfer reaction catalyzed by dihydrofolate reductase. *Biochemistry*, 42(46):13558–13575, 2003.

[28] P. T. R. Rajagopalan, S. Lutz, and S. J. Benkovic. Coupling interactions of distal residues enhance dihydrofolate reductase catalysis: Mutational effects on hydride transfer rates. *Biochemistry*, 41(42):12618–12628, 2002.

[29] E. Neria and M. Karplus. Molecular dynamics of an enzyme reaction: Proton transfer in TIM. *Chem. Phys. Lett.*, 267(1-2):23–30, 1997.

[30] J. Z. Pu, S. H. Ma, M. Garcia-Viloca, J. L. Gao, D. G. Truhlar, and A. Kohen. Nonperfect synchronization of reaction center rehybridization in the transition state of the hydride transfer catalyzed by dihydrofolate reductase. *J. Am. Chem. Soc.*, 127(42):14879–14886, 2005.

[31] Q. Cui and M. Karplus. Quantum mechanics/molecular mechanics studies of triosephosphate isomerase-catalyzed reactions: Effect of geometry and tunneling on proton-transfer rate constants. *J. Am. Chem. Soc.*, 124(12):3093–3124, 2002.

[32] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory. Comput.*, 4(3):435–447, 2008.

[33] W. E and E. Vanden-Eijnden. Towards a theory of transition paths. *J. Stat. Phys.*, 123(3):503–523, 2006.

Chapter 2

Quantum Simulation of Temperature Independent Kinetic Isotope Effects in Enzyme Catalysis

A temperature independent kinetic isotope effect (KIE) of hydrogen transfer is feature of many wild type enzymes functioning at their optimal conditions on their natural substrates, yet this phenomenon is not predicted by simple reaction rate models. Here, we use ring polymer molecular dynamics to demonstrate temperature independent KIEs in the dihydrofolate reductase system and reveal that compression of the hydride donor and acceptor atoms during the reaction event leads to transient configurations for hydride transfer that are mass independent.

2.1 Introduction

A broad range of biologically relevant reactions involve the transfer of a hydrogen, either as a hydrogen atom, a proton, or a hydride. Studies of the kinetic isotope effect (KIE)—the decrease in the intrinsic reaction rate constant when a light isotope is substituted for a heavy one—of such reactions holds much promise in understanding the catalytic efficiency of enzymes and how it is influenced by factors such as quantum mechanical tunneling, protein motion, and mutation. A growing number of KIE studies reveal a feature of enzyme catalyzed hydrogen

transfer: wild-type enzymes functioning on their natural substrate in their optimal temperature range exhibit temperature independent KIEs. Examples of such temperature independence are summarized in Table 2.1. This temperature independence challenges interpretation of reaction rates based on rate models that simply include zero point energy and tunneling corrections [1].

Model rate theories that predict temperature independent KIEs have been developed [2, 3, 4]. The mechanistic interpretation of such temperature independence is that organization of the enzyme preceeds hydrogen transfer to an extent that is independent of hydrogen mass. This organization gives rise to an Arrhenius activation energy that is independent of mass; the magnitude of the KIE is due to mass dependent differences in tunneling from the ground state. Alternatively, simulations based on the variational transition state theory (VTST) method [5] have shown that temperature independent KIEs arise due to a temperature dependent change in the classical dividing surface position. In these simulations, the Arrhenius activation energy is mass independent due to changes in the quantum effects of barrier crossing with temperature which cancel most of mass dependent difference in activation energy due to zero-point energy.

We explore these two mechanisms using ring polymer molecular dynamics (RPMD) [6] to simulate hydride transfer catalyzed by *Escherichia coli* dihydrofolate reductase. RMPD is a path-integral method that enables inclusion of nuclear quantization effects, such as the zero-point energy and tunneling, in the dynamics of the transferring hydride and yields reaction rates and mechanisms that are formally independent of the choice of dividing surface (Figure 2.1) or any other reaction coordinate assumption . The RPMD method enables generation of the ensemble of reactive, quantized molecular dynamics trajectories, which is essential for the following analysis of dynamical correlations. In this current study, RPMD is used to directly simulate hydride transfer catalyzed by dihydrofolate reductase over a range of temperatures with the three isotopes of hydrogen. These simulations, which reproduce the weak temperature dependence of experimental kinetic isotope effects of DHFR catalysis, allow for the analysis of the RPMD reactive

Enzyme	H/D KIE	Temperature range (K)
DHFR	3.5	278-318
tDHFR	4	298-338
TSOX	7	278-308
TMADH	7	278-298
MADH	17	278-313
AcCoA-DS	23	278-303
ThyS	7	278-313
SLO	81	283-323
tADH	3.2	303-338

Table 2.1: Enzymes exhibiting temperature independent KIEs on their natural substrate and conditions. DHFR, dihydrofolate reductase [7]; tDHFR, thermophilic alcohol dehydrogenase [8]; TSOX, sarcosine oxidase [9]; TMADH, trimethylamine dehydrogenase [10]; MADH, methylamine dehydrogenase [3]; AcCoA-DS, AcCoA desaturase [11]; ThyS, thymidylate synthase [12, 13]; SLO, soybean lipoxygenase [14]; tADH, thermophilic alcohol dehydrogenase [15, 16].

trajectories to elucidate the mechanistic origins of temperature independent KIEs.

2.2 Methods

2.2.1 Ring polymer molecular dynamics

The RPMD equations of motion for a system with a quantized hydride of mass m and N classical particles are [17, 18]

$$\dot{\mathbf{v}}^{(\alpha)} = \omega_n^2 \left(\mathbf{q}^{(\alpha+1)} + \mathbf{q}^{(\alpha-1)} - 2\mathbf{q}^{(\alpha)} \right) - \frac{1}{m} \nabla_{\mathbf{q}^{(\alpha)}} U_{\text{ext}} \left(\mathbf{q}^{(\alpha)}, \mathbf{Q}_1, \ldots, \mathbf{Q}_N \right) \tag{2.1}$$

and

$$\dot{\mathbf{V}}_j = -\frac{1}{nM_j} \sum_{\alpha=1}^{n} \nabla_{\mathbf{Q}_j} U_{\text{ext}} \left(\mathbf{q}^{(\alpha)}, \mathbf{Q}_1, \ldots, \mathbf{Q}_N \right), \tag{2.2}$$

where $\mathbf{q}^{(\alpha)}$ and $\mathbf{v}^{(\alpha)}$ are the position and velocity vectors of the α^{th} ring polymer bead, $\mathbf{Q}_j$ and $\mathbf{V}_j$ are the position and velocity of the j^{th} classical particle with cor-

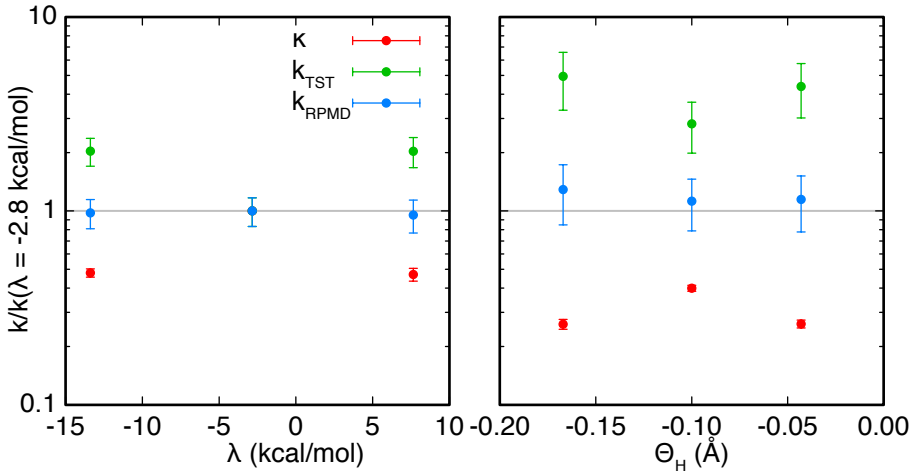

Figure 2.1: RPMD rates are independent of the choice of dividing surface. Here, the rate of hydride transfer catalyzed by DHFR is calculated using two collective variable definitions of the dividing surface, λ and Θ_{H}, described in the text, at three values for each coordinate. The ratio with respect to the dividing surface $\lambda = -2.8$ kcal/mol is shown for each term in the RPMD rate—the transmission coefficient κ (red) and the TST rate (green)–and the RPMD rate (blue). The independence of the RPMD rate is achieved as the larger TST rate is offset by a higher degree of recrossing.

responding mass M_j, and n is the number of imaginary time ring-polymer beads. The intra-bead harmonic frequency is $\omega_n = \frac{n}{\beta\hbar}$ where β is the reciprocal temperature. $U_{\mathrm{ext}}\left(\mathbf{q}^{(\alpha)}, \mathbf{Q}_1, \ldots, \mathbf{Q}_N\right)$ is the potential energy function of the system, and $\mathbf{q}^{(0)} = \mathbf{q}^{(n)}$. Equations 2.1 and 2.2 generate classical dynamics that we employ as a model for the real-time dynamics of the system [19]. In the limit of large n, these time-reversible dynamics preserve the exact Boltzmann distribution [20, 21, 22].

As in classical formulations of the thermal rate constant [23, 24, 25], the RPMD rate equation can be written as [26, 27]

$$k_{\mathrm{RPMD}} = \lim_{t\to\infty} \kappa\left(t\right) k_{\mathrm{TST}}. \tag{2.3}$$

Here, k_{TST} is the transition state theory (TST) approximation for the rate for a diving surface $\lambda\left(\mathbf{x}\right) = \lambda^{\ddagger}$, where vector $\mathbf{x}$ is the position the system including the positions of beads of the quantized hydride and the positions of all classical

nuclei. The prefactor κ is the transmission coefficient that accounts for recrossing of trajectories through the dividing surface. As is the case for both exact classical and exact quantum dynamics, the RPMD method yields reaction rates and mechanisms that are independent of the choice of the dividing surface [26, 27, 28].

The TST rate in Equation 2.3 is calculated using [29, 30, 31, 32]

$$k_{\mathrm{TST}} = (2\pi\beta)^{-1/2} \langle g_\lambda \rangle_\ddagger \frac{e^{-\beta \Delta F(\lambda^\ddagger)}}{\int_{-\infty}^{\lambda^\ddagger} d\lambda e^{-\beta \Delta F(\lambda)}}. \tag{2.4}$$

Here, $F(\lambda)$ is the free energy (FE) along λ such that

$$e^{-\beta \Delta F(\lambda')} = \frac{\langle \delta(\lambda(\mathbf{x}) - \lambda') \rangle}{\langle \delta(\lambda(\mathbf{x}) - \lambda_r) \rangle}, \tag{2.5}$$

where λ_r is a reference point in the reactant region, and [29, 33]

$$g_\lambda(\mathbf{x}) = \left[\sum_{i=1}^{d} \frac{1}{m_i} \left(\frac{\partial \lambda(\mathbf{x})}{\partial x_i} \right)^2 \right]^{1/2}, \tag{2.6}$$

where $x_i \in \{\mathbf{x}\}$ indicates either a ring-polymer bead or classical particle degree of freedom, m_i is the corresponding mass, and d is the total number of degrees of freedom in the system. In Equation 2.5, $\langle \cdots \rangle$ denotes the equilibrium ensemble average

$$\langle \cdots \rangle = \frac{\int d\mathbf{x} \int d\mathbf{v}\, e^{-\beta H_n(\mathbf{x},\mathbf{v})} (\cdots)}{\int d\mathbf{x} \int d\mathbf{v}\, e^{-\beta H_n(\mathbf{x},\mathbf{v})}}, \tag{2.7}$$

and $\langle \cdots \rangle_\ddagger$ in Equation 2.4 denotes the average in the transition-state constrained ensemble

$$\langle \cdots \rangle_\ddagger = \frac{\int d\mathbf{x} \int d\mathbf{v}\, e^{-\beta H_n(\mathbf{x},\mathbf{v})} (\cdots) \delta\left(\lambda(\mathbf{x}) - \lambda^\ddagger\right)}{\int d\mathbf{x} \int d\mathbf{v}\, e^{-\beta H_n(\mathbf{x},\mathbf{v})} \delta(\lambda(\mathbf{x}) - \lambda^\ddagger)}. \tag{2.8}$$

Here,

$$H_n(\mathbf{x},\mathbf{v}) = \sum_{j=1}^{N} \frac{1}{2} M_j \mathbf{V}_j^2 + \sum_{\alpha=1}^{n} \frac{1}{2} m_{\mathrm{b}} \left(\mathbf{v}^{(\alpha)} \right)^2 + U_n(\mathbf{x}), \tag{2.9}$$

where m_{b} is the Parrinello-Rahman mass [21], $\mathbf{v} = \left\{ \mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(n)}, \mathbf{V}_1, \ldots, \mathbf{V}_N \right\}$,

and

$$U_n(\mathbf{x}) = \frac{1}{n}\sum_{\alpha=1}^{n}\frac{1}{2}m\omega_n^2\left(q^{(\alpha)} - q^{(\alpha-1)}\right)^2 + \frac{1}{n}\sum_{\alpha=1}^{n}U_{\text{ext}}(\mathbf{x}^{(\alpha)}) \tag{2.10}$$

is the full potential energy function for the RPMD system where $\mathbf{x}^{(\alpha)} \equiv (\mathbf{q}^{(\alpha)}, \mathbf{Q}_1, \ldots, \mathbf{Q}_N)$ is used to indicate the position of ring polymer bead α and the full set of classical nuclei.

The transmission coefficient in Equation 2.3 is obtained from the flux-side correlation function [26, 27],

$$\kappa(t) = \frac{\left\langle \dot{\lambda}_0 h\left(\lambda(\mathbf{x}_t) - \lambda^{\ddagger}\right)\right\rangle_{\ddagger}}{\left\langle \dot{\lambda}_0 h\left(\dot{\lambda}_0\right)\right\rangle_{\ddagger}} \tag{2.11}$$

by releasing trajectories from the transition-state constrained ensemble. Here, h is the Heaviside function, $\dot{\lambda}_0$ is the time derivative of λ upon initialization of the RPMD trajectory from the dividing surface with initial velocities sampled from the Maxwell-Boltzmann distribution, and $\mathbf{x}_t$ is the time-evolved position of the system along the RPMD trajectory.

2.2.2 Calculation details

RPMD simulations of DHFR were performed using a modified version of the Gromacs-4.5.5 molecular dynamics package [34] at three temperatures spanning the range of experimental KIEs: 280, 300, and 320K. At each temperature, the reaction was simulated with the mass of the transferring hydride at each of the three isotopes of hydrogen.

The full potential energy function for the RPMD system is described using the empirical valence bond (EVB) method [35, 36],

$$\begin{aligned} U_{\text{ext}}(\mathbf{x}^{(\alpha)}) &= \frac{1}{2}\left(V_{\text{r}}(\mathbf{x}^{(\alpha)}) + V_{\text{p}}(\mathbf{x}^{(\alpha)})\right) \\ &- \frac{1}{2}\sqrt{\left(V_{\text{r}}(\mathbf{x}^{(\alpha)}) - V_{\text{p}}(\mathbf{x}^{(\alpha)})\right)^2 + 4V_{12}^2}. \end{aligned} \tag{2.12}$$

The terms $V_{\text{r}}(\mathbf{x}^{(\alpha)})$ and $V_{\text{p}}(\mathbf{x}^{(\alpha)})$ are the molecular mechanics potential energy

functions for the system with the hydride covalently bonded to the donor and acceptor atoms based on previously described [37] modified version of the GROMOS 43A1 united atom forcefield [38], with modifications described in Appendix B. The constant V_{12} =13.4 kcal/mol is fit to the Eyring equation for the experimental rate of the intrinsic reaction [39], and the product state potential $V_{\rm p}(\mathbf{x}^{(j)})$ includes a constant shift of $\Delta_{12} = +101.9$ kcal/mol to match the experimental driving force for the intrinsic reaction [39].

System preparation

The system is initialized and equilibrated at each temperature following a previously described protocol [37]. The system is initialized from the DHFR crystal structure in the active configuration (PDB code: 1RX2) [40]. Crystallographic 2-mercaptoethanol and manganese ions are removed; crystallographic waters are not. The amine side chain of Q102 is rotated 180 degrees to correctly coordinate the adenine moiety of the cofactor [41]. To be consistent with the observed hydrogen bonding networks in the crystal structure, histidine residues 45, 124, and 149 are protonated at nitrogen ND1, histidine residues 114 and 141 are protonated at nitrogen NE2, and both DHFR cysteine residues are in their protonated form [36]. The enzyme is explicitly solvated using 4,122 SPC/E rigid water molecules [42] in a truncated octahedral simulation cell with constant volume and periodic boundary conditions. The periodic image distance for the cell is 57.686 Å. Twelve Na^+ ions are included for charge neutrality. The full system includes N =14,080 classical nuclei.

From the initial geometry of the crystal structure, the system is equilibrated on the reactant potential energy surface $V_{\rm r}$ using classical MD. In a series of three equilibration steps, MD trajectories of length 10 ps in time are performed with progressively weaker harmonic restraints between the heavy atom positions and the crystal structure; the restraint force constants for the three equilibration runs are 100, 50, and 25 kcal mol^{-1}Å^{-2}, respectively, and the runs are performed in the NVT ensemble using Berendsen thermostat with a coupling constant of 0.01 ps

[43]. After initial equilibration to the reactants basin, the system is equilibrated on the full potential energy surface (Equation A.1) for an additional 100 ps of classical MD. Finally, the ring polymer representation for the quantized hydride was introduced at the geometry of the relaxed classical system and equilibrated for an additional 1 ps using RPMD with velocities resampled from the Maxwell-Boltzmann distribution every 100 fs.

RPMD simulations

In all simulations, the RPMD equations of motion are evolved using the velocity Verlet algorithm [44]. As in previous RPMD simulations, each timestep involves separate coordinate updates for forces arising from the external potential and for exact evolution of the purely harmonic portion of the ring-polymer potentials [6]. The hydride is quantized with $n = 32$ beads for each of the three masses.

Reaction progress is characterized by the collective variable λ $\lambda \equiv \lambda(\mathbf{x}^{\mathrm{c}}) = V_{\mathrm{r}}(\mathbf{x}^{\mathrm{c}}) - V_{\mathrm{p}}(\mathbf{x}^{\mathrm{c}})$, where $\mathbf{x}^{\mathrm{c}} \equiv (\mathbf{q}^{\mathrm{c}}, \mathbf{Q}_1, \ldots, \mathbf{Q}_N)$, $\mathbf{q}^{\mathrm{c}} = \sum_{\alpha=1}^{n} \mathbf{q}^{(\alpha)}/n$ is the ring polymer centroid with mass $m_{\mathrm{c}} = m$, and V_{r} and V_{p} are defined above.

The RPMD reaction rate is calculated from the product of the TST rate and the transmission coefficient (Equation 2.3). The FE profiles that appear in the TST rate expression (Equation 2.4) are obtained using umbrella sampling along λ and the weighted histogram analysis method (WHAM) [45]. The umbrella sampling method [46] is used to efficiently sample this collective variable between the reactant and product basins. Independent RPMD sampling trajectories are performed using biased potentials of the form

$$U_n(\mathbf{x}) + \frac{1}{2} k_l \left(\lambda(\mathbf{x}^{\mathrm{c}}) - \lambda_l\right)^2, \qquad l = 1, \ldots, 25, \tag{2.13}$$

where the $\{k_l\}$ and $\{\lambda_l\}$ are listed in Table B.1.

For the RPMD trajectories used to sample the FE profile, to diminish the separation of timescales for the motion of the ring polymer and the rest of the system the mass of the ring polymer centroid is m =12 amu, and the masses of the

harmonic internal modes of the ring polymer are scaled so each mode has a period of 8 fs. Changing these parameters does not affect the ensemble of configurations that are sampled in the calculation of the FE profile; it merely allows for the sampling trajectories to be performed with a larger simulation time-step (1.0 fs) than is used in the dynamical trajectories. Furthermore, unlike the RPMD dynamical trajectories in which the long-range electrostatic contributions are updated every time-step, we use twin-ranged cut-offs [38] in the FE sampling trajectories such that non-bonding interactions beyond 9 Å are updated every 5 fs. Sampling trajectories are performed at constant temperature by resampling the particle velocities from the Maxwell-Boltzmann distribution every 1.3 ps.

For each temperature and mass, the sampling trajectories are initialized in order of increasing λ_l, as follows. The first sampling trajectory ($l = 1$) was initialized from the equilibrated system in the reactant basin. After 25 ps of simulation, the configuration from this first sampling trajectory was used to initialize the second sampling trajectory ($l = 2$). After 25 ps of simulation, the configuration from the second trajectory was used to initialize the third sampling trajectory ($l = 3$), and so on. After initialization, each window was run for 2 ns. The full initialization protocol was repeated five times for each window giving a total of 10 ns sampled for each value of l. In order to better converge the k_{TST} term, windows between reactants and products were repeated an additional 5 times, giving a total of 20 ns sampled for windows of $l < 18$. The weighted histogram analysis method (WHAM) [45] is used to calculate the unbiased FE profile $F(\lambda)$ from the set of sampling trajectories.

Equilibrium averages

Equilibrium ensemble averages are presented for the system in the reactant region, the dividing surface region, and the product region. These ensemble averages are

strictly defined using

$$\begin{aligned}\langle A\rangle_{\lambda^*} &= \int_{\lambda^*-\delta\lambda}^{\lambda^*+\delta\lambda} d\lambda' P(\lambda') \int d\mathbf{q}_1 \ldots \int d\mathbf{q}_n \int d\mathbf{Q}_1 \ldots \int d\mathbf{Q}_N \\ &\times \delta(\lambda' - \lambda(\mathbf{x}))A(\mathbf{x})),\end{aligned} \tag{2.14}$$

where $P(\lambda) = \exp(-\beta F(\lambda))/\int d\lambda' \exp(-\beta F(\lambda'))$, and $F(\lambda)$ is calculated using umbrella sampling, as described above. Similarly, equilibrium ensemble distributions are presented for each of the three regions. These distributions are defined using

$$\begin{aligned}P_{\lambda^*}(A) &= \int_{\lambda^*-\delta\lambda}^{\lambda^*+\delta\lambda} d\lambda' P(\lambda') \int d\mathbf{q}_1 \ldots \int d\mathbf{q}_n \int d\mathbf{Q}_1 \ldots \int d\mathbf{Q}_N \\ &\times \delta(\lambda' - \lambda(\mathbf{x}))\delta(A - A(\mathbf{x}))).\end{aligned} \tag{2.15}$$

For the ensembles in the reactant, dividing surface, and product regions, we employ λ^* = -141 kcal/mol, -2.5 kcal/mol, and 162 kcal/mol, respectively, and $\delta\lambda$ = 2.5 kcal/mol.

The dividing surface ensemble

Boltzmann-weighted sampling on the reaction dividing surface is performed with constrained molecular dynamics using the RATTLE algorithm [47]. The existing implementation of RATTLE in Gromacs-4.5.5 is modified to constrain both classical MD and RPMD with respect to the collective variable $\lambda(\mathbf{x}^c)$. To remove the hard-constraint bias from the ensemble of configurations that is sampled in the constrained dynamics [48, 49], each sampled configuration is weighted by $[H(\mathbf{x}^c)]^{-1/2}$, where

$$H(\mathbf{x}^c) = (m_c)^{-1} \left|\nabla_{\mathbf{q}^c}\lambda(\mathbf{x}^c)\right|^2 + \sum_{k=1}^{N} M_k^{-1} \left|\nabla_{\mathbf{Q}_k}\lambda(\mathbf{x}^c)\right|^2. \tag{2.16}$$

For each temperature and mass, 1008 independent RPMD trajectories are run with the dividing surface constraint. These constrained trajectories are initialized from configurations near the dividing surface obtained in the umbrella sampling

trajectories that are restrained to the dividing surface region using Equation 2.13, and they are performed at constant temperature by resampling velocities from the Maxwell-Boltzmann distribution every 1.3 ps. Following an initial equilibration of 10 ps, each of the constrained trajectories is run for 20 ps, and dividing surface configurations are sampled every 4 ps. As with the umbrella sampling trajectories, the constrained dynamics are run with a timestep of 1 fs enabled by scaling masses of the harmonic internal modes of the ring polymer so each mode has a period of 8 fs.

RPMD transition path ensemble

As we have done previously [37, 50], we analyze the transition path ensemble [51] for the RPMD trajectories in this study. Reactive trajectories are generating by forward and backward integration of initial configurations drawn from the dividing surface ensemble with initial velocities drawn from the MB distribution. Reactive trajectories correspond to those for which the forward and backward integrated half-trajectories terminate on opposite sides of the diving surface. The reactive trajectory, being initialized on the dividing surface must be reweighed to obtain the unbiased transition path ensemble [32, 52]. A weighting term w_α is applied to each trajectory, correctly accounting for recording and for the fact that individual trajectories are performed in the microcanonical ensemble. This term is given by

$$w_\alpha = \left(\sum_i \left| \dot{\lambda} \left(\mathbf{x} \right) \right|^{-1} \right)^{-1}, \tag{2.17}$$

where the sum includes all instances in which trajectory α crosses the diving surface, and $\dot{\lambda} \left(\mathbf{x} \right)$ is the velocity of the collective variable at recrossing event i. The reweighhting has a minor effect on the non-equilibrium averages if the reactive trajectories initialized from the dividing surface exhibit few recrossings, as is the case here. Non-equilibrium averages over the RPMD transition path ensemble are calculated by aligning reactive trajectories at time 0, defined as the moment in time when the trajectories are released from the dividing surface.

T (K)	Mass	$\lambda^{\ddagger}$ (kcal/mol)	k_{TST} (s^{-1})	κ	k_{RPMD} (s^{-1})
280	H	-2.6	1100 ± 100	0.808 ± 0.004	900 ± 100
280	D	-2.1	200 ± 20	0.872 ± 0.007	170 ± 20
280	T	-2.0	110 ± 10	0.881 ± 0.006	100 ± 10
300	H	-2.8	3800 ± 400	0.826 ± 0.008	3100 ± 400
300	D	-2.2	750 ± 80	0.874 ± 0.006	660 ± 80
300	T	-2.1	360 ± 20	0.884 ± 0.005	320 ± 20
320	H	-2.9	5400 ± 100	0.832 ± 0.008	4500 ± 100
320	D	-2.2	1200 ± 100	0.877 ± 0.005	1100 ± 100
320	T	-2.1	710 ± 80	0.888 ± 0.003	630 ± 80

Table 2.2: Hydride transfer rates and terms for each temperature and mass.

2.3 Results

2.3.1 RPMD rates and KIEs

Figure 2.1 demonstrates the independence of the RPMD rate with regards to mechanistic choices of reaction coordinate and dividing surface. The rate of hydride transfer was calculated using three values of the energy gap coordinate λ (-13.4 kcal/mol, -2.8 kcal/mol, and 7.6 kcal/mol) and three values of the hydride coordinate Θ_{H} (-0.168 Å, -0.100 Å, and -0.043 Å) which describes motion of the hydride between the donor and acceptor carbons using

$$\Theta_{\mathrm{H}} = |\mathbf{Q}_{\mathrm{D}} - \mathbf{q}^{\mathrm{c}}| - |\mathbf{Q}_{\mathrm{A}} - \mathbf{q}^{\mathrm{c}}| \tag{2.18}$$

where $\mathbf{Q}_{\mathrm{D}}$ and $\mathbf{Q}_{\mathrm{A}}$ are the position vectors for the donor and acceptor carbons, respectively. The transition state theory rate is overestimated by shifting the dividing surface away from λ = -2.8 kcal/mol or by defining the dividing surface on the more-local hydride coordinate. Figure 2.1 demonstrates that overestimation is compensated for by the transmission coefficient providing the dividing surface independence of the RPMD rate.

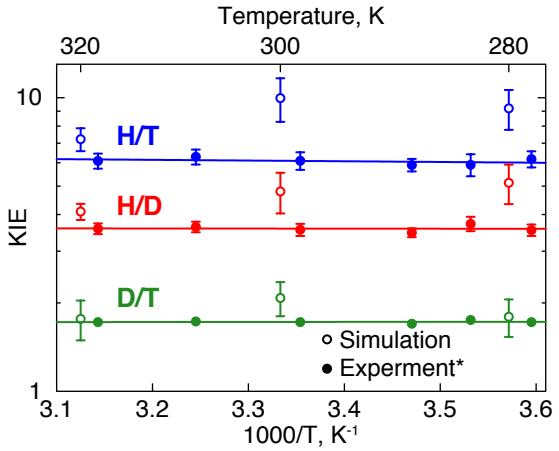

Figure 2.2: Simulated (open circles) and experimental KIEs for the intrinsic reaction. *—Experimental KIEs from [7]

Kinetic isotope effects for the RPMD system are calculated using $(L/\Gamma)\mathrm{KIE} = k^{L}_{\mathrm{RPMD}}/k^{\Gamma}_{\mathrm{RPMD}}$ where L/Γ is H/D, H/T or D/T. These KIEs qualitatively reproduce the magnitudes and small temperature dependence of experimental KIEs (Figure 2.2). Here, the dividing surface for each temperature and mass was determined by the maximum value of λ on the FE profile from a harmonic fit in the transition state region. These values ($\lambda^{\ddagger}$), as well as the rates (k_{RPMD} and k_{TST}) and the transmission coefficient (κ) are presented for each temperature and mass in Table 2.2. Arrhenius plots for the calculated KIEs are shown with experimental intrinsic reaction rates [7].

2.3.2 Mass dependence of organization of the enzyme

The minimum distance of approach between the hydride-transfer donor and acceptor atoms is used to report on enzyme organization accompanying hydride transfer.

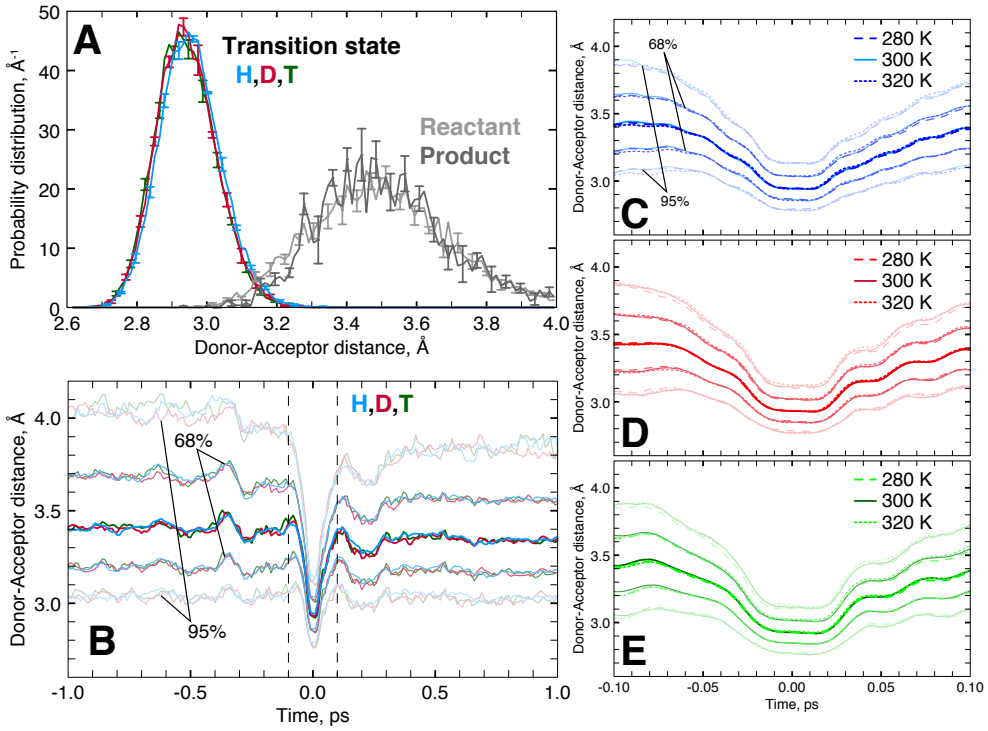

Figure 2.3: Donor-acceptor distance in the equilibrium ensemble and during the intrinsic reaction (*A*) The equilibrium probability distributions of the donor-acceptor distance is shown in the reactant (light grey), the product (dark grey), and the transition state (blue) regions for the hydrogen reaction at 300 K. The equilibrium probability distributions in the transition state region are shown for deuterium (red) and tritium (green). (*B*) The time-dependent distribution of donor-acceptor distances in the non-equilibrium ensemble of reactive trajectories for each mass at 300 K is shown for hydrogen (blue), deuterium (red) and tritium (blue). The median value of these distributions is in bold, and contours representing 68% and 95% are shown. (*C*-*E*) The time-dependent distribution of donor-acceptor distances in the non-equilibrium ensemble of reactive trajectories is shown from -0.1 to 0.1 ps (vertical dashed lines in *B*) for each of the three temperatures 280 K (dashed), 300 K (solid), and 320K (dotted) for each of the masses, hydrogen (*C*), deuterium (*D*), and tritium (*E*).

Donor-acceptor distance sampling is considered in the equilibrium ensemble and the ensemble of reactive trajectories. Figure 2.3A shows the Boltzmann weighted equilibrium distribution of donor-acceptor distances in the reactant, product and transition state ensembles. The reaction proceeds via compression from 3.5 Å to 2.9 Å in the donor-acceptor distance coordinate. The distribution in the transition state is remarkably independent of hydride mass. The nature of how this

compression is sampled is explored by looking at the distribution of donor acceptor distances in the ensemble of reactive trajectories. The narrow distribution at the transition state quickly relaxes to distributions consistent with the reactant and product basins within 150 fs. Similar to the equilibrium distributions, the distribution in the ensemble of reactive trajectories is mass independent and temperature independent (2.3C-E).

2.3.3 Temperature dependence of classical barrier position

To compare to previous VTST studies [5], The temperature dependent shift in classical barrier observed in previous VTSTS studies was estimated by considering the temperature dependence of classical barrier position on the same coordinate, Θ_H (Equation 2.18) was estimated by considering the free energy profile for tritium transfer. Figure 2.4 shows a shift of 0.013 $\pm$ 0.002 Å on Θ_H towards products for the system at 320 K compared to 280 K.

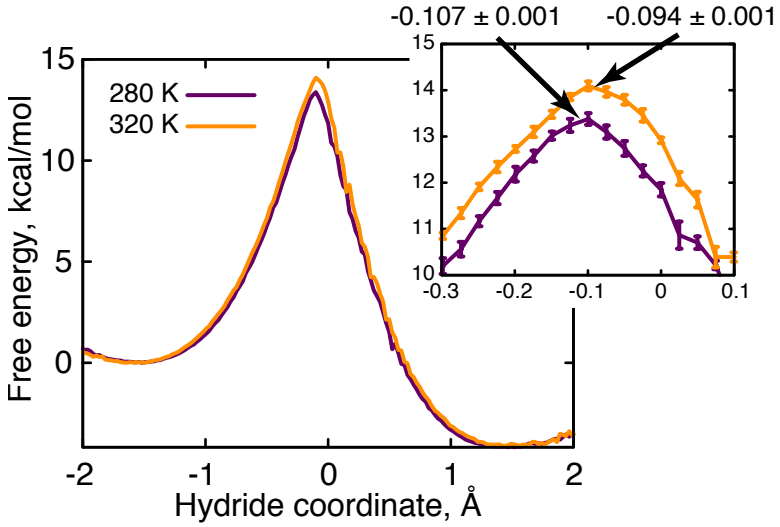

Figure 2.4: The temperature dependent shift in classical reaction barrier position on the hydride coordinate Θ_H is estimated by considering the FE profile on Θ_H for tritium transfer. At 320 K, the FE barrier is shifted towards products in comparison to 280 K (inset). Barrier position (arrows) is determined using a harmonic fit to the barrier region; errors are estimated by fitting five data blocks.

2.3.4 Dynamical correlations

Dynamical correlations in the intrinsic reaction are characterized following previously developed measures of dynamical correlation [37]. A measure of velocity cross-correlations in the reactive trajectories is introduced, $d_{ij}(t) = D_{ij}(t)/(\langle D_{ii}\rangle \langle D_{jj}\rangle)^{1/2}$ such that

$$D_{ij}(t) = \langle \mathbf{v}_i \cdot \mathbf{v}_j \rangle_t \tag{2.19}$$

and

$$\langle D_{ii}\rangle = \langle \mathbf{v}_i \cdot \mathbf{v}_i \rangle . \tag{2.20}$$

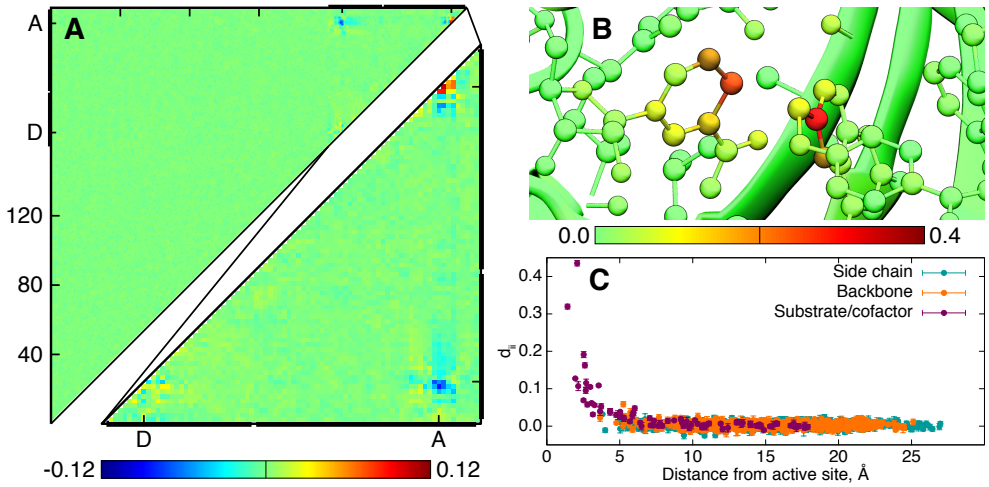

Figure 2.5: Dynamical correlations among enzyme motions during the intrinsic reaction involving hydrogen at 300K. (*A*) (Upper triangle) The integrated dynamical correlation measure d_{ij}. Protein residues are indexed according to PDB 1RX2; substrate and cofactor regions are indicated by the hydride acceptor A and donor D atoms, respectively. Significant dynamical correlations appear primarily in the substrate and cofactor regions, which are enlarged in the lower triangle. (*D*) The color of atoms in the active site region are scaled according to the diagonal of the integrated dynamical correlation measure, d_{ii}. (*C*) The diagonal of the integrated dynamical correlation measure, d_{ii}, as a function of the distance of atom i from the midpoint of the donor and acceptor atoms. Atoms corresponding to the protein side chains, the protein backbone, and the substrate/cofactor regions are indicated by color.

Here, $\langle ... \rangle_t$ denotes an average over the non-equilibrium ensemble of phase-space points that lie on reactive trajectories which crossed the dividing surface some time t earlier and subsequently terminate in the product basin, and $\langle ... \rangle$ denotes the equilibrium average. $d_{ij}(t)$, which vanishes for the equilibrium ensemble, reports on the degree to which atoms move in concert during the intrinsic reaction step.

Figure 2.5 summarizes the extent of dynamical correlations throughout the enzyme system in terms of $d_{ij} = \int_{-\tau}^{\tau} d_{ij}(t)dt$. Negative dynamical correlations are seen between the donor and acceptor carbon atoms (Figure 2.5A), which move in opposite directions (first approaching each other, then moving apart) during the hydride transfer. Similarly, positive correlations are seen between atom pairs on the cofactor and on the substrate which move in concert as the hydride is transferred. Figure 2.5A is reproduced for all masses and temperatures in Figure B.1 Diagonal terms d_{ii} report on the degree to which the velocity of an individual atom deviates from equilibrium during the transient reaction event. Fig 2.5B shows the active site color coded by the value of the diagonal term d_{ii} for each atom. Figure 2.5C correlates the diagonal terms with distance from the active site. Previous results are reconfirmed: Atom pairs that that are dynamically correlated and atoms that undergo significant deviations from their equilibrium behavior are highly localized to the active site [37].

Transition state tunneling

Tunneling through the transition state along the donor-acceptor coordinate is characterized using the connection between RPMD and semiclassical instanton theory [53, 54] and recognizing that an extended configuration of the ring polymer at the dividing surface indicates that the system is in a tunneling regime. The radius of gyration

$$R_g^2 = \frac{1}{n} \sum_{\alpha=1}^{n} \left| \mathbf{q}^{(\alpha)} - \mathbf{q}^{\mathrm{c}} \right|^2 \tag{2.21}$$

is used to characterize this extension. We define a measure of extension along the donor-acceptor coordinate as

$$\left(R_g^{\mathrm{DA}}\right)^2 = \frac{1}{n}\sum_{\alpha=1}^{n}\left(q_{\mathrm{DA}}^{(\alpha)} - q_{\mathrm{DA}}^{\mathrm{c}}\right)^2, \tag{2.22}$$

where $q_{\mathrm{DA}}^{(\alpha)}$ is the scalar projection of the α^{th} bead onto the vector connecting the donor and acceptor carbons. To characterize extension or compression transverse to the donor-acceptor coordinate, we define

$$\left(R_g^{\perp}\right)^2 = \frac{R_g^2 - \left(R_g^{\mathrm{DA}}\right)^2}{2}. \tag{2.23}$$

We consider

$$r_g^{\mathrm{DA}}(t) = \frac{3\left\langle \left(R_g^{\mathrm{DA}}\right)^2 \right\rangle_t}{\left\langle R_g^2 \right\rangle_{\lambda_{\mathrm{R}}}} \tag{2.24}$$

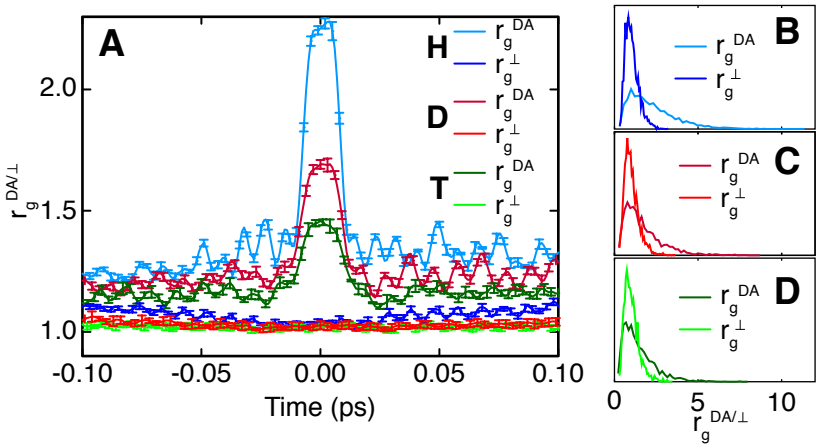

Figure 2.6: Nuclear quantum effects at the transition state for each of the three masses at 300 K are characterized by considering transition state tunneling along donor-acceptor bond and constriction in transverse coordinates using the radius of gyration based measures r_g^{DA} and $r_g^{\perp}$. (A) The time evolution of r_g^{DA} and $r_g^{\perp}$ in the ensemble of reactive trajectories. (B-D) The distribution of r_g^{DA} and $r_g^{\perp}$ in the ensemble of reactive trajectories at t=0 for hydrogen (A), deuterium (B), and tritium (C).

and

$$r_g^{\perp}(t) = \frac{3\left\langle \left(R_g^{\perp}\right)^2 \right\rangle_t}{\left\langle R_g^2 \right\rangle_{\lambda_{\mathrm{R}}}} \tag{2.25}$$

to characterize deviations from the equilibrium ensemble of these values in the ensemble of reactive trajectories. Here, λ_{R} is the average over the reactants basin.

Fig 2.6A-D shows significant extension of the ring polymer along the donor-acceptor coordinate during the intrinsic reaction event and quick collapse away from the barrier (Fig 2.6A). These deviations increase with decreasing mass. There is no extension or compression in coordinates transverse to the donor-acceptor coordinate, suggesting that constriction of the hydrogen in these coordinates is unlikely to play a role in determining the KIE.

2.4 Conclusion

Quantized molecular dynamics simulations qualitatively reproduce the experimentally observed small temperature dependence of the KIE of hydride transfer catalyzed by dihydrofolate reductase (Figure 2.2). These simulations reveal that organization of the enzyme prior to hydride transfer gives rise to compressed donor-acceptor distances that are independent of mass (Figure 2.3), and that this compression gates tunneling in hydride transfer (Figure 2.6). These results confirm the mechanism predicted by model rate equations [2, 3, 4]. These simulations also reveal a temperature dependent shift in the classical barrier to hydride transfer (Figure 2.4) confirming the mechanism of previous simulation studies based on the VTST method [5]. Taken together, these results reconcile these two mechanisms by revealing that they are not necessarily mutually exclusive. This study demonstrates the strength of RPMD as a quantum simulation method that enables the calculation of reaction rates without prior assumptions of the underlying mechanism.

References

[1] R. P. Bell. *The Tunnel Effect in Chemistry*. Chapman and Hall, New York, 1980.

[2] M. J. Knapp and J. P. Klinman. Environmentally coupled hydrogen tunneling—linking catalysis to dynamics. *Eur. J. Biochem.*, 269(13):3113–3121, 2002.

[3] J. Basran, M. J. Sutcliffe, and N. S. Scrutton. Enzymatic H-transfer requires vibration-driven extreme tunneling. *Biochemistry*, 38(10):3218–3222, 1999.

[4] A. M. Kuznetsov and J. Ulstrup. Proton and hydrogen atom tunnelling in hydrolytic and redox enzyme catalysis. *Can. J. Chem.*, 77(5-6):1085–1096, 1999.

[5] J. Z. Pu, S. H. Ma, J. L. Gao, and D. G. Truhlar. Small temperature dependence of the kinetic isotope effect for the hydride transfer reaction catalyzed by *Escherichia coli* dihydrofolate reductase. *J. Phys. Chem. B*, 109(18):8551–8556, 2005.

[6] S. Habershon, D. E. Manolopoulos, T. E. Markland, and T. F. Miller III. Ring-polymer molecular dynamics: Quantum effects in chemical dynamics from classical trajectories in an extended phase space. *Annu. Rev. Phys. Chem.*, 64:387–413, 2013.

[7] R. S. Sikorski, L. Wang, K. A. Markham, P. T. R. Rajagopalan, S. J. Benkovic, and A. Kohen. Tunneling and coupled motion in the *Escherichia coli* dihydrofolate reductase catalysis. *J. Am. Chem. Soc.*, 126(15):4778–4779, 2004.

[8] G. Maglia and R. K. Allemann. Evidence for environmentally coupled hydrogen tunneling during dihydrofolate reductase catalysis. *J. Am. Chem. Soc.*, 125(44):13372–13373, 2003.

[9] R. J. Harris, R. Meskys, M. J. Sutcliffe, and N. S. Scrutton. Kinetic studies of the mechanism of carbon-hydrogen bond breakage by the heterotetrameric sarcosine oxidase of *Arthrobacter* sp. 1-IN. *Biochemistry*, 39(6):1189–1198, 2000.

[10] J. Basran, M. J. Sutcliffe, and N. S. Scrutton. Deuterium isotope effects during carbon-hydrogen bond cleavage by trimethylamine dehydrogenase—implications for mechanism and vibrationally assisted hydrogen tunneling in wild-type and mutant enzymes. *J. Biol. Chem.*, 276(27):24581–24587, 2001.

[11] J. L. Abad, F. Camps, and G. Fabrias. Is hydrogen tunneling involved in Acyl-CoA desaturase reactions? the case of a δ^9 desaturase that transforms (*E*)-11-tetradecenoic acid into (*Z*,*E*)-9,11-tetradecadienoic acid. *Angew. Chem. Int. Ed.*, 39(18):3279–3281, 2000.

[12] N. Agrawal, B. Y. Hong, C. Mihai, and A. Kohen. Vibrationally enhanced hydrogen tunneling in the *Escherichia coli* thymidylate synthase catalyzed reaction. *Biochemistry*, 43(7):1998–2006, 2004.

[13] B. Hong, F. Maley, and A. Kohen. Role of Y94 in proton and hydride transfers catalyzed by thymidylate synthase. *Biochemistry*, 46(49):14188–14197, 2007.

[14] M. J. Knapp, K. Rickert, and J. P. Klinman. Temperature-dependent isotope effects in soybean lipoxygenase-1: Correlating hydrogen tunneling with protein dynamics. *J. Am. Chem. Soc.*, 124(15):3865–3874, 2002.

[15] A. Kohen, R. Cannio, S. Bartolucci, and J. P. Klinman. Enzyme dynamics and hydrogen tunnelling in a thermophilic alcohol dehydrogenase. *Nature*, 399(6735):496–499, 1999.

[16] Z. X. Liang, T. Lee, K. A. Resing, N. G. Ahn, and J. P. Klinman. Thermal-activated protein mobility and its correlation with catalysis in thermophilic alcohol dehydrogenase. *Proc. Natl. Acad. Sci. U.S.A.*, 101(26):9556–9561, 2004.

[17] I. R. Craig and D. E. Manolopoulos. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *J. Chem. Phys.*, 121(8):3368–3373, 2004.

[18] Richard P. Feynman and Albert R. Hibbs. *Quantum mechanics and path integrals*. International series in pure and applied physics. McGraw-Hill, New York, 1965.

[19] T. F. Miller III. Isomorphic classical molecular dynamics model for an excess electron in a supercritical fluid. *J. Chem. Phys.*, 129(19), 2008.

[20] D. Chandler and P. G. Wolynes. Exploiting the isomorphism between quantum-theory and classical statistical-mechanics of polyatomic fluids. *J. Chem. Phys.*, 74(7):4078–4095, 1981.

[21] M. Parrinello and A. Rahman. Study of an F-center in molten KCl. *J. Chem. Phys.*, 80(2):860–867, 1984.

[22] B. De Raedt, M. Sprik, and M. L. Klein. Computer-simulation of muonium in water. *J. Chem. Phys.*, 80(11):5719–5724, 1984.

[23] E. Wigner. Concerning the excess of potential barriers in chemical reactions. *Zeitschrift Fur Physikalische Chemie-Abteilung B-Chemie Der Elementarprozesse Aufbau Der Materie*, 19(2/3):203–216, 1932.

[24] H. Eyring. The activated complex in chemical reactions. *J. Chem. Phys.*, 3(2):107–115, 1935.

[25] J. C. Keck. Variational theory of chemical reaction rates applied to 3-body recombinations. *J. Chem. Phys.*, 32(4):1035–1050, 1960.

[26] I. R. Craig and D. E. Manolopoulos. A refined ring polymer molecular dynamics theory of chemical reaction rates. *J. Chem. Phys.*, 123(3):034102, 2005.

[27] I. R. Craig and D. E. Manolopoulos. Chemical reaction rates from ring polymer molecular dynamics. *J. Chem. Phys.*, 122(8):084106, 2005.

[28] W. H. Miller. Improved classical path approximation for boltzmann density matrix. *J. Chem. Phys.*, 58(4):1664–1667, 1973.

[29] R. Collepardo-Guevara, I. R. Craig, and D. E. Manolopoulos. Proton transfer in a polar solvent from ring polymer reaction rate theory. *J. Chem. Phys.*, 128(14):144502, 2008.

[30] A. R. Menzeleev, N. Ananth, and T. F. Miller III. Direct simulation of electron transfer using ring polymer molecular dynamics: Comparison with semiclassical instanton theory and exact quantum methods. *J. Chem. Phys.*, 135(7), 2011.

[31] D. Chandler. Statistical-mechanics of isomerization dynamics in liquids and transition-state approximation. *J. Chem. Phys.*, 68(6):2959–2970, 1978.

[32] H. Bennett Charles. Molecular dynamics and transition state theory: The simulation of infrequent events. In *Algorithms for Chemical Computations*, volume 46 of *ACS Symposium Series*, pages 63–97. American Chemical Society, 1977. doi:10.1021/bk-1977-0046.ch004.

[33] G. K. Schenter, B. C. Garrett, and D. G. Truhlar. Generalized transition state theory in terms of the potential of mean force. *J. Chem. Phys.*, 119(12):5828–5833, 2003.

[34] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory. Comput.*, 4(3):435–447, 2008.

[35] A. Warshel and R. M. Weiss. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Am. Chem. Soc.*, 102(20):6218–6226, 1980.

[36] P. K. Agarwal, S. R. Billeter, and S. Hammes-Schiffer. Nuclear quantum effects and enzyme dynamics in dihydrofolate reductase catalysis. *J. Phys. Chem. B*, 106(12):3283–3293, 2002.

[37] N. Boekelheide, R. Salomon-Ferrer, and T. F. Miller III. Dynamics and dissipation in enzyme catalysis. *Proc. Natl. Acad. Sci. U.S.A.*, 108(39):16159–16163, 2011.

[38] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger Mark, A. E., W. R. P. Scott, and I. G. Tironi. *Biomolecular Simulation: The GROMOS96 manual and user guide*. Hochschuleverlag AG an der ETH Zürich, Zürich, Switzerland, 1996.

[39] C. A. Fierke, K. A. Johnson, and S. J. Benkovic. Construction and evaluation of the kinetic scheme associated with dihydrofolate-reductase from *Escherichia coli*. *Biochemistry*, 26(13):4085–4092, 1987.

[40] M. R. Sawaya and J. Kraut. Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: Crystallographic evidence. *Biochemistry*, 36(3):586–603, 1997.

[41] I. V. Khavrutskii, D. J. Price, J. Lee, and C. L. Brooks. Conformational change of the methionine 20 loop of *Escherichia coli* dihydrofolate reductase modulates pk(a) of the bound dihydrofolate. *Protein Sci.*, 16(6):1087–1100, 2007.

[42] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987.

[43] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and

J. R. Haak. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[44] L. Verlet. Computer experiments on classical fluids I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159(1):98–103, 1967.

[45] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.*, 13(8):1011–1021, 1992.

[46] G. M. Torrie and J. P. Valleau. Non-physical sampling distributions in Monte-Carlo free-energy estimation—Umbrella sampling. *J. Comput. Phys.*, 23(2):187–199, 1977.

[47] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical-integration of cartesian equations of motion of a system with constraints—Molecular-dynamics of N-alkanes. *J. Comput. Phys.*, 23(3):327–341, 1977.

[48] D. A. Zichi, G. Ciccotti, J. T. Hynes, and M. Ferrario. Molecular-dynamics simulation of electron-transfer reactions in solution. *J. Phys. Chem.*, 93(17):6261–6265, 1989.

[49] D. Frenkel and B. Smit. *Understanding molecular simulation: From algorithms to applications*. Academic Press, San Diego, 2nd edition, 2002.

[50] J. S. Kretchmer and T. F. Miller III. Direct simulation of proton-coupled electron transfer across multiple regimes. *The Journal of chemical physics*, 138(13):134109–134109, 2013.

[51] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.*, 53:291–318, 2002.

[52] G. Hummer. From transition paths to transition states and rate coefficients. *J. Chem. Phys.*, 120(2):516–523, 2004.

[53] J. O. Richardson and S. C. Althorpe. Ring-polymer instanton method for calculating tunneling splittings. *J. Chem. Phys.*, 134(5), 2011.

[54] S. C. Althorpe. On the equivalence of two commonly used forms of semiclassical instanton theory. *J. Chem. Phys.*, 134(11), 2011.

Chapter 3

Promoting Modes in Lactate Dehydrogenase

3.1 Introduction

Lactate dehydrogenase (LDH) catalyzes the reversible transfer of a hydride and proton to pyruvate to produce lactate. This reaction controls levels of cellular pyruvate when oxygen is in limited supply. Isozymes of LDH exhibit different kinetic parameters that fine-tune the regulation of cellular pyruvate to meet the particular metabolic needs of a cell. Understanding the mechanistic origin of the differences in kinetic properties between LDH isozymes motivated one of the first applications of transition path sampling (TPS) in enzyme catalysis [1, 2, 3, 4]. Intriguingly, the ensemble of reactive trajectories determined in these studies revealed spatially extensive, sub-picosecond motion in the protein that couples to active site compression during the reaction event (Figure 3.1).

We will explore the nature of this compressive mode in the context of dynamics and dissipation during the transient reaction event [5] by determining the degree to which this mode appears in dynamical and statistical correlations along the reaction path. Here, we use RPMD free energy sampling trajectories to explore the degree to which the potential energy surface used to describe the active site determines the reaction mechanism and energetics [6].

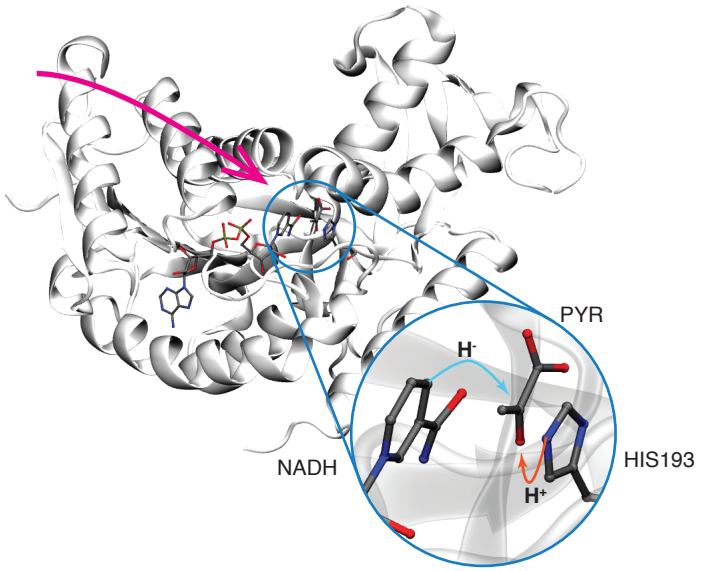

Figure 3.1: The crystal structure of human heart lactate dehydrogenase isozyme which catalyzes the conversion of pyruvate to lactate by transferring a proton from hisidine 193 and a hydride from NAPH to pyruvate (inset). The reported compressive mode [3] includes residues under the magenta arrow. (Protein Data Bank structure 1I0Z [7]).

3.2 Methods

3.2.1 Ring polymer molecular dynamics

The RPMD equations of motion for the LDH holoenzyme with a quantized proton and a hydride of masses m and with N classical particles are [8, 9]

$$\dot{\mathbf{v}}_{\mathrm{P}}^{(\alpha)} = \omega_n^2 \left(\mathbf{q}_{\mathrm{P}}^{(\alpha+1)} + \mathbf{q}_{\mathrm{P}}^{(\alpha-1)} - 2\mathbf{q}_{\mathrm{P}}^{(\alpha)} \right) - \frac{1}{m} \nabla_{\mathbf{q}_{\mathrm{P}}^{(\alpha)}} U_{\mathrm{ext}} \left(\mathbf{q}_{\mathrm{P}}^{(\alpha)}, \mathbf{q}_{\mathrm{H}}^{(\alpha)}, \mathbf{Q}_1, \ldots, \mathbf{Q}_N \right), \tag{3.1}$$

$$\dot{\mathbf{v}}_{\mathrm{H}}^{(\alpha)} = \omega_n^2 \left(\mathbf{q}_{\mathrm{H}}^{(\alpha+1)} + \mathbf{q}_{\mathrm{H}}^{(\alpha-1)} - 2\mathbf{q}_{\mathrm{H}}^{(\alpha)} \right) - \frac{1}{m} \nabla_{\mathbf{q}_{\mathrm{H}}^{(\alpha)}} U_{\mathrm{ext}} \left(\mathbf{q}_{\mathrm{P}}^{(\alpha)}, \mathbf{q}_{\mathrm{H}}^{(\alpha)}, \mathbf{Q}_1, \ldots, \mathbf{Q}_N \right), \tag{3.2}$$

and

$$\dot{\mathbf{V}}_j = -\frac{1}{nM_j} \sum_{\alpha=1}^{n} \nabla_{\mathbf{Q}_j} U_{\mathrm{ext}} \left(\mathbf{q}_{\mathrm{P}}^{(\alpha)}, \mathbf{q}_{\mathrm{H}}^{(\alpha)}, \mathbf{Q}_1, \ldots, \mathbf{Q}_N \right), \tag{3.3}$$

where $\mathbf{q}_{\mathrm{P}}^{(\alpha)}$, $\mathbf{v}_{\mathrm{P}}^{(\alpha)}$, $\mathbf{q}_{\mathrm{H}}^{(\alpha)}$ and $\mathbf{v}_{\mathrm{H}}^{(\alpha)}$ are the position and velocity vectors of the α^{th} proton (P) and hydride (H) ring polymer bead, $\mathbf{Q}_j$ and $\mathbf{V}_j$ are the position and

velocity of the j^{th} classical particle with corresponding mass M_j, and n is the number of imaginary time ring-polymer beads. The intra-bead harmonic frequency is $\omega_n = \frac{n}{\beta\hbar}$ where β is the reciprocal temperature. $U_{\text{ext}}\left(\mathbf{q}_{\text{P}}^{(\alpha)}, \mathbf{q}_{\text{H}}^{(\alpha)}, \mathbf{Q}_1, \ldots, \mathbf{Q}_N\right)$ is the potential energy function of the system, $\mathbf{q}_{\text{P}}^{(0)} = \mathbf{q}_{\text{P}}^{(n)}$, and $\mathbf{q}_{\text{H}}^{(0)} = \mathbf{q}_{\text{H}}^{(n)}$. Equations 3.1, 3.2 and 3.3 generate classical dynamics that we employ as a model for the real-time dynamics of the system [10]. In the limit of large n, these dynamics preserve the exact Boltzmann distribution [11, 12, 13].

3.2.2 Calculation details

Quantum mechanical/classical mechanics potential energy function

A quantum mechanical/classical mechanical (QM/MM) potential energy function [14] is used to describe U_{ext} appearing in Equations 3.1, 3.2 and 3.3. The QM/MM potential is obtained by dividing the system into two regions: A quantum mechanical (QM) subregion whose potential is calculated using a QM Hamiltonian, and a molecular mechanical (MM) region system comprised of the remaining atoms whose potential is calculated using a MM force field. Interactions across the two regions are treated using the electrostatic embedding method [14] with link-atom hydrogens capping the covalent bonds that cross the QM/MM boundary [15]. In the LDH system, the QM region (Figure 3.2) is composed of 39 atoms in the active site including the nicotinamide ring of NADH, the pyruvate molecule, the imiazole ring and β-carbon of histidine-193, and two link-atom hydrogens attached to the N_1-nitrogen in the nicotinamide ring and the β-carbon of histidine 193. The CHARMM27 MM force field [16] with CMAP dihedral cross-term corrections [17] is used for the MM region. Two semi-empirical QM methods are used for the QM region: The AM1 Hamiltonian [18] and a semi-empirical method based on the AM1 Hamiltonian reparameterized to reproduce higher level electronic structure energies for the QM region in the LDH system. This reparameterization is described below.

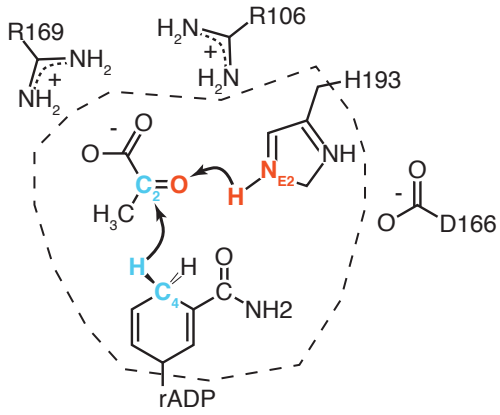

Figure 3.2: The QM region of the LDH system is made up of 39 atoms contained in the dashed line. Active site residues which stabilize the pyruvate and histidine groups are shown.

System preparation

The system is initialized from the crystal structure of human heart L-LDH monomer complexed with NADH and oxamate [7] (Protein Data Bank structure 1I0Z). Crystalographic waters are kept in the structure, and the active-site bound oxamate molecule is made a pyruvate molecule by changing the amino group to a methyl group. Following previous studies [2], the dimer-binding loop (residues 1-19) is omitted to reduce the simulation cell size. Based on observed hydrogen bonding networks in the crystal structure, histidine residue 156 is protonated at the ND1-nitrogen, histidine residues 67, 181, 186, 231, and 271 are protonated at the NE2-nitrogen, and active-site histidine residue 193 is protonated at both ND1- and NE2-nitrogens. The enzyme is explicitly solvated using 16,931 TIP3P water molecules [19] with Lennard-Jones parameters on water hydrogens following CHARMM convention [16] in a truncated octahedral simulation cell with constant volume and periodic boundary conditions. The periodic image distance for the cell is 90Å. Six Na^+ ions are included for charge neutrality. The full system includes 55,766 classical nuclei.

From the initial geometry of the solvated crystal structure, the system is equilibrated on an MM potential energy surface with the QM region frozen. Non-bonded

interactions between the MM region and the frozen QM region in this equilibration potential energy surface are determined by assigning charges and Lennard-Jones terms to the atoms in the QM region from analogous atoms in the CHARMM27 force field. In a series of three equilibration steps, MD trajectories of length 10 ps in time are performed with progressively weaker harmonic restraints between the heavy atom positions and the crystal structure; the restraint force constants for the three equilibration runs are 100, 50, and 25 kcal mol^{-1} Å^{-2}, respectively, and the runs are performed in the NVT ensemble using a Berendsen thermostat with coupling constant of 0.01 ps [20]. After initial equilibration with the QM region frozen, the system is equilibrated on the full QM/MM potential energy surface using the AM1 Hamiltonian for the QM region for an additional 100 ps of classical MD. Finally, the ring polymer representation of the quantized proton and hydride is introduced and equilibrated for an additional 1 ps using RPMD velocities RPMD with velocities resampled from the Maxwell-Boltzmann distribution every 0.1 ps.

RPMD simulations

The RPMD equations of motion are evolved using the velocity Verlet algorithm [21]. As in previous RPMD simulations, each timestep involves separate coordinate updates for forces arising from the external potential and for exact evolution of the purely harmonic portion of the ring-polymer potentials [22]. The proton and hydride are quantized with $n = 32$ beads.

Proton and hydride transfer are characterized by the collective variables Θ_{P} and Θ_{H}, respectively, which are defined as

$$\Theta_{\mathrm{P}} = |\mathbf{Q}_{\mathrm{D_P}} - \mathbf{q}_{\mathrm{P}}^{\mathrm{c}}| - |\mathbf{Q}_{\mathrm{A_P}} - \mathbf{q}_{\mathrm{P}}^{\mathrm{c}}| \tag{3.4}$$

and

$$\Theta_{\mathrm{H}} = |\mathbf{Q}_{\mathrm{D_H}} - \mathbf{q}_{\mathrm{H}}^{\mathrm{c}}| - |\mathbf{Q}_{\mathrm{A_H}} - \mathbf{q}_{\mathrm{H}}^{\mathrm{c}}|\,, \tag{3.5}$$

where $\mathbf{Q}_{\mathrm{D_P}}$ and $\mathbf{Q}_{\mathrm{A_P}}$ are the position vectors of the proton donor and acceptor atoms, respectively, and $\mathbf{q}_{\mathrm{P}}^{\mathrm{c}} = \sum_{\alpha=1}^{n} \mathbf{q}_{\mathrm{P}}^{(\alpha)}/n$ is the centroid of the proton ring

polymer, and $\mathbf{Q}_{\mathrm{D_H}}$, $\mathbf{Q}_{\mathrm{A_H}}$, and $\mathbf{q}^{\mathrm{c}}_{\mathrm{H}}$ are similarly defined for the hydride.

Two-dimensional free energy profiles on Θ_{P} and Θ_{H} are obtained using umbrella sampling [23] and the weighted histogram analysis method (WHAM) [24]. The umbrella-sampling biasing potentials have the form

$$\frac{1}{2}k\left(\Theta_{\mathrm{P}}\left(\mathbf{x}^{\mathrm{c}}\right)-\Theta_{l_{\mathrm{P}}}\right)^2+\frac{1}{2}k\left(\Theta_{\mathrm{H}}\left(\mathbf{x}^{\mathrm{c}}\right)-\Theta_{l_{\mathrm{H}}}\right)^2, \qquad l_{\mathrm{P}}=1,\ldots,14; l_{\mathrm{H}}=1,\ldots,14, \tag{3.6}$$

where the force constant k has a value of 180 kcal mol^{-1}Å^{-2}, and the $\Theta_{l_{\mathrm{P}}}$ and $\Theta_{l_{\mathrm{H}}}$ are both evenly spaced from -1.3 to 1.3 Å.

For the RPMD trajectories used to sample the FE profile, to diminish the separation of timescales for the motion of the ring polymer and the rest of the system the mass of the ring polymer centroid is m =12 amu, and the masses of the harmonic internal modes of the ring polymer are scaled so each mode has a period of 8 fs. Changing these parameters does not affect the ensemble of configurations that are sampled in the calculation of the FE profile; it merely allows for the sampling trajectories to be performed with a larger simulation time-step (0.001 ps) than would be possible using physical masses. Furthermore, the long-range electrostatic contributions are updated every time-step, and we use twin-ranged cut-offs [25] in the FE sampling trajectories such that non-bonding interactions beyond 10 Å are updated every 5 fs. Sampling trajectories are performed at constant temperature by resampling the particle velocities from the Maxwell-Boltzmann distribution every 1.3 ps.

Umbrella sampling trajectories were initialized from the reactants in order of increasing l_{P} and l_{H} as follows. The first sampling trajectory (l_{P} = 1, l_{H} = 1) was initialized from the RPMD equilibrated system in the reactant basin. After 5 ps of simulation, the configuration from this first sampling trajectory was used to initialize the next set of sampling trajectories (l_{P} = 1, l_{H} = 2), (l_{P} = 2, l_{H} = 1), and (l_{P} = 2, l_{H} = 2). After 5 ps of simulation, the configurations from this set were used to initialize the third set of sampling trajectories and so on. After initialization, each window was run for 500 ps for a total of 95 ns simulation time.

WHAM was used to calculate the unbiased 2D FE surface from this set of sampling trajectories.

Specific reaction parameter semi-empirical Hamiltonian

A specific reaction parameter (SRP) Hamiltonian was developed. Following the SRP approach [26], the QM region is treated by a semi-empirical model that is specifically parameterized to reproduce DFT or *ab initio* energies. Specifically, the parameter set of the AM1 Hamiltonian is optimized such that the resultant semi-empirical QM/MM potential reproduces the QM/MM potential with the energy of the QM region calculated using second order Møller-Plesset perturbation theory (MP2) [27] using the 6-31G** basis set [28]. The choice of this higher order method and basis set is based on previously reported results [6].

MP2 QM/MM calculations were done using the QM/MM implementation in Gromacs-4.5.5 [29] interfaced with the ORCA 2.9 electronic structure package [30]. Semi-empiricla QM/MM calculations were done using the QM/MM implementation in Gromacs-4.5.5 [29] interfaced with the MOPAC semi-empirical quantum chemistry package [31]

MP2 calculations were done on a set of configurations of the LDH system generated as follows: An initial configuration was selected from configurations generated using FE sampling described above. From this initial configuration a set of minimized configurations was generated by freezing atoms further than 10 Å away from the QM region, and constraining Θ_P and Θ_H to a set of specific values. Minimization was done on the AM1 QM/MM potential energy surface using steepest descent. The constrained values of Θ_P and Θ_H were selected to grid the configuration space spanned by Θ_P and Θ_H. Specifically, these values were evenly spaced by 0.3 Å from -1.5 to 1.5 Å with additional grid points at $\pm$1.8 Å with the other coordinate spanning -0.9 to 0.9 Å. The largest discrepancy between the AM1 and PM2 energies for the QM subsystem is observed in the box spanned by -0.5 to 0.5 Å in each coordinate [6], so 0.1 Å grid was used in this box. This method gives a total of 261 grid points for each initial configuration selected

from FE sampling. Seven initial configurations were used to give a total of 1827 configurations parameter optimization.

In order to better capture the specific interactions of bond making and breaking that are required for the reaction, the AM1 parameter set was expanded. Four new parameter sets were added such that the AM1 parameters for each donor and acceptor atom were uniquely optimized. The expanded AM1 parameter set includes 164 free parameters. A sequential constrained optimization scheme was performed following a previously described SRP protocol [32]. (1) The one-center energies (U_{ss} and U_{pp}) and resonance integrals (β_s and β_p) were optimized. (2) Next, the α parameters in the core-repulsion function were added, followed by (3) the orbital exponents (ζ_s and ζ_p), then (4) the one-center two-electron repulsion integrals (G_{ss}, G_{sp}, G_{pp}, $G_{pp'}$ and H_{sp}), and finally the full set of parameters including the Gaussian core repulsion function parameters L, M, and K. To avoid large changes in parameter values during optimization, constraints were put on the parameters as follows: For atoms H, C, N and O, all parameters were constrained to be within 15% the of the original AM1 value. For the additional donor and acceptor atom parameters, all parameters except L, M, and K were constrained to be within 20% of the original, and L, M, and K were allowed to vary up to 25%. The constrained optimization was done using the non-linear least squares solver included in the Intel MKL library.

3.3 Results and Discussion

The two dimensional FE surface on Θ_{P} and Θ_{H} calculated using the AM1 semi-empirical QM/MM potential energy surface surface shows two reaction channels– the H^+/H^- channel with H^+-transfer followed by H^--transfer and the H^-/H^+ channel with H^--transfer followed by H^+-transfer. The H^+/H^- channel slightly favored, but both channels give a barrier over 50 kcal/mol. The reaction on this potential is endergonic by 30 kcal/mol. These values are inconsistent with experimental measurements with a barrier of $\sim$14 kcal/mol [33, 34, 35] and near-zero

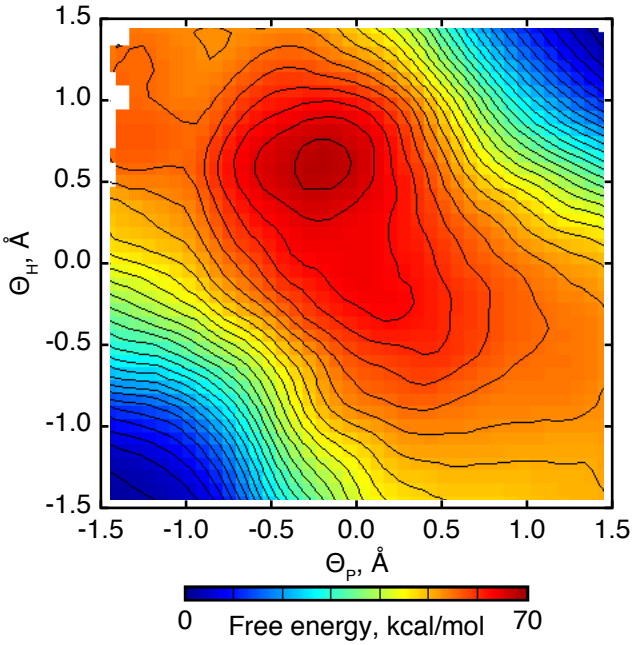

Figure 3.3: The FE surface on Θ_{P} and Θ_{H} calculated using the AM1 potential.

driving force.

The semi-empirical AM1-SRP potential reproduces MP2 energies with a standard deviation of 1 kcal/mol in contrast to the 12 kcal/mol of the AM1 potential 3.4. The two dimensional FE surface on Θ_{P} and Θ_{H} calculated using the AM1-SRP semi-empirical QM/MM potential energy surface (Figure 3.5). This FE surface shows a single concerted reaction channel, reproducing the mechanistic dependence on QM potential previously reported using AM1 and MP2 potential energy surfaces [6].

One dimensional free energy sampling was done along this channel on the coordinate $\Theta_{\mathrm{HP}} = \frac{1}{2}(\Theta_{\mathrm{P}} + \Theta_{\mathrm{H}})$. 1D sampling trajectories were initialized from configurations from the 2D FE sampling trajectories along the line $\Theta_{\mathrm{P}} = \Theta_{\mathrm{H}}$. Twenty-five umbrella sampling trajectories with 1D biasing potentials Table C.2 were run for 1 ns each, and WHAM was used to calculate the unbiased 1D FE profile (Figure 3.6). The FE barrier on the AM1-SRP potential energy surface is 16 kcal/mole, and the driving force is 1.6 kcal/mol.

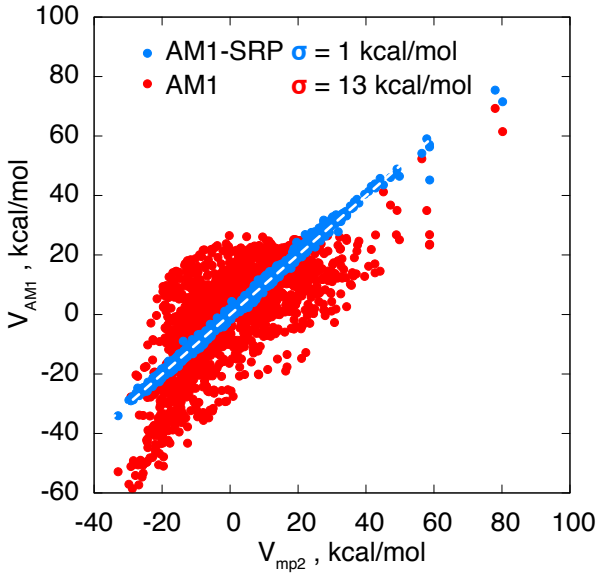

Figure 3.4: AM1 and AM1-SRP energies of the training set used for SRP optimization are plotted versus MP2 energies. The standard deviation of the AM1-SRP energy from the MP2 energy is 1 kcal/mol.

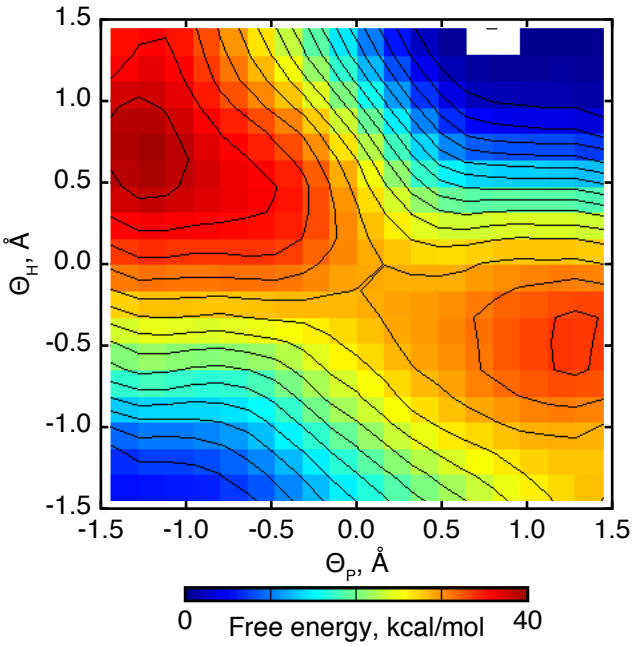

Figure 3.5: The FE surface on Θ_P and Θ_H calculated using the AM1-SRP potential.

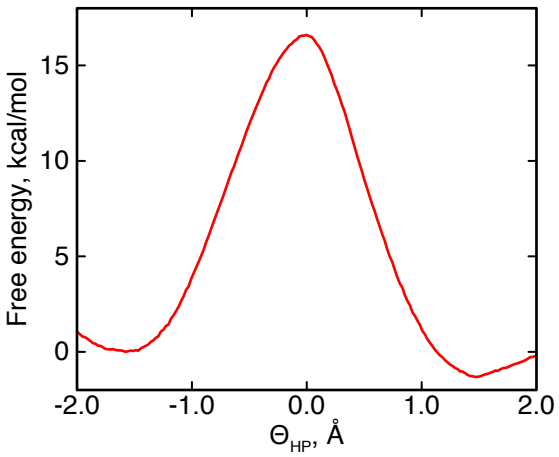

Figure 3.6: The 1D FE profile along Θ_{HP}.

3.4 Conclusion

The semi-empirical AM1-SRP QM/MM potential provides forces and energies that are consistent with the higher level MP2 QM/MM potential it was optimized to reproduce yet is fast enough to sample dynamical trajectories. The FE barrier for the reaction on this potential is 16 kcal/mol which compares favorably with experimental rates–more favorably than the barrier of 50 kcal/mol calculated using the AM1 potential, which has been used in previous studies [2, 3, 4]. This study has provided a potential on which to explore the nature of the correlation between protein motions and motions in the active site.

References

[1] S. D. Schwartz and V. L. Schramm. Enzymatic transition states and dynamic motion in barrier crossing. *Nat Chem Biol*, 5(8):552–559, 2009.

[2] J. E. Basner and S. D. Schwartz. Donor-acceptor distance and protein promoting vibration coupling to hydride transfer: A possible mechanism for kinetic control in isozymes of human lactate dehydrogenase. *J. Phys. Chem. B*, 108(1):444–451, 2004.

[3] S. L. Quaytman and S. D. Schwartz. Reaction coordinate of an enzymatic reaction revealed by transition path sampling. *Proc. Natl. Acad. Sci. U.S.A.*, 104(30):12253–12258, 2007.

[4] A. Davarifar, D. Antoniou, and S. D. Schwartz. The promoting vibration in human heart lactate dehydrogenase is a preferred vibrational channel. *J. Phys. Chem. B*, 115(51):15439–15444, 2011.

[5] N. Boekelheide, R. Salomon-Ferrer, and T. F. Miller III. Dynamics and dissipation in enzyme catalysis. *Proc. Natl. Acad. Sci. U.S.A.*, 108(39):16159–16163, 2011.

[6] S. Ferrer, E. Silla, I. Tunon, M. Oliva, V. Moliner, and I. H. Williams. Dependence of enzyme reaction mechanism on protonation state of titratable residues and QM level description: Lactate dehydrogenase. *Chem Commun (Cambridge, U K)*, (47):5873–5875, 2005.

[7] J. A. Read, V. J. Winter, C. M. Eszes, R. B. Sessions, and R. L. Brady. Structural basis for altered activity of M- and H-isozyme forms of human

lactate dehydrogenase. *Proteins-Structure Function and Genetics*, 43(2):175–185, 2001.

[8] I. R. Craig and D. E. Manolopoulos. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *J. Chem. Phys.*, 121(8):3368–3373, 2004.

[9] Richard P. Feynman and Albert R. Hibbs. *Quantum mechanics and path integrals*. International series in pure and applied physics. McGraw-Hill, New York, 1965.

[10] T. F. Miller III. Isomorphic classical molecular dynamics model for an excess electron in a supercritical fluid. *J. Chem. Phys.*, 129(19), 2008.

[11] D. Chandler and P. G. Wolynes. Exploiting the isomorphism between quantum-theory and classical statistical-mechanics of polyatomic fluids. *J. Chem. Phys.*, 74(7):4078–4095, 1981.

[12] M. Parrinello and A. Rahman. Study of an F-center in molten KCl. *J. Chem. Phys.*, 80(2):860–867, 1984.

[13] B. De Raedt, M. Sprik, and M. L. Klein. Computer-simulation of muonium in water. *J. Chem. Phys.*, 80(11):5719–5724, 1984.

[14] M. J. Field, P. A. Bash, and M. Karplus. A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations. *J. Comput. Chem.*, 11(6):700–733, 1990.

[15] U. C. Singh and P. A. Kollman. A combined *ab initio* quantum-mechanical and molecular mechanical method for carrying out simulations on complex molecular-systems—applications to the $CH_3Cl + Cl^-$ exchange-reaction and gas-phase protonation of polyethers. *J. Comput. Chem.*, 7(6):718–730, 1986.

[16] A. D. Mackerell, N. Banavali, and N. Foloppe. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, 56(4):257–265, 2001.

[17] A. D. Mackerell, M. Feig, and C. L. Brooks. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.*, 25(11):1400–1415, 2004.

[18] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. The development and use of quantum-mechanical molecular-models .76. AM1—a new general-purpose quantum-mechanical molecular-model. *J. Am. Chem. Soc.*, 107(13):3902–3909, 1985.

[19] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, 1983.

[20] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and J. R. Haak. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[21] L. Verlet. Computer experiments on classical fluids I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159(1):98–103, 1967.

[22] S. Habershon, D. E. Manolopoulos, T. E. Markland, and T. F. Miller III. Ring-polymer molecular dynamics: Quantum effects in chemical dynamics from classical trajectories in an extended phase space. *Annu. Rev. Phys. Chem.*, 64:387–413, 2013.

[23] G. M. Torrie and J. P. Valleau. Non-physical sampling distributions in Monte-Carlo free-energy estimation—Umbrella sampling. *J. Comput. Phys.*, 23(2):187–199, 1977.

[24] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.*, 13(8):1011–1021, 1992.

[25] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger Mark, A. E., W. R. P. Scott, and I. G. Tironi. *Biomolecular Simulation: The GROMOS96 manual and user guide*. Hochschuleverlag AG an der ETH Zürich, Zürich, Switzerland, 1996.

[26] I. Rossi and D. G. Truhlar. Parameterization of NDDO wave-functions using genetic algorithms—An evolutionary approach to parameterizing potential-energy surfaces and direct dynamics calculations for organic-reactions. *Chem. Phys. Lett.*, 233(3):231–236, 1995.

[27] M. Head-Gordon, J. A. Pople, and M. J. Frisch. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.*, 153(6):503–506, 1988.

[28] G. A. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Allaham, W. A. Shirley, and J. Mantzaris. A complete basis set model chemistry. 1. The total energies of closed-shell atoms and hydrides of the 1st-row elements. *J. Chem. Phys.*, 89(4):2193–2218, 1988.

[29] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory. Comput.*, 4(3):435–447, 2008.

[30] F Neese. *ORCA—An* Ab initio*, Density Functional and Semiempirical Program Package, version 2.9.1*. Department of molecular theory and spectroscopy, Max Planck Institute for Bioinorganic Chemistry, Mulheim/Ruhr, Germany, 1996.

[31] M.J.S Dewar. Development and status of MINDO/3 and MNDO. *J. Mol. Struct.*, 100(41), 1983.

[32] D. Doron, D. T. Major, A. Kohen, W. Thiel, and X. Wu. Hybrid quantum and classical simulations of the dihydrofolate reductase catalyzed hydride transfer reaction on an accurate semi-empirical potential energy surface. *J. Chem. Theory. Comput.*, 7(10):3420–3437, 2011.

[33] N. Zhadin, M. Gulotta, and R. Callender. Probing the role of dynamics in hydride transfer catalyzed by lactate dehydrogenase. *Biophys J*, 95(4):1974–1984, 2008.

[34] A. R. Clarke, D. B. Wigley, W. N. Chia, D. Barstow, T. Atkinson, and J. J. Holbrook. Site-directed mutagenesis reveals role of mobile arginine residue in lactate-dehydrogenase catalysis. *Nature*, 324(6098):699–702, 1986.

[35] D. K. Shoemark, M. J. Cliff, R. B. Sessions, and A. R. Clarke. Enzymatic properties of the lactate dehydrogenase enzyme from *Plasmodium falciparum*. *FEBS J*, 274(11):2738–2748, 2007.

Appendix A

Appendix 1: Supplemental Information for Chapter 1

A.1 Potential Energy Surface

The potential energy surface for the hydride transfer reaction in DHFR is described using the EVB method [1, 2],

$$\begin{aligned} U(\mathbf{x}^{(j)}) &= \frac{1}{2}\left(V_{\mathrm{r}}(\mathbf{x}^{(j)}) + V_{\mathrm{p}}(\mathbf{x}^{(j)})\right) \\ &- \frac{1}{2}\sqrt{\left(V_{\mathrm{r}}(\mathbf{x}^{(j)}) - V_{\mathrm{p}}(\mathbf{x}^{(j)})\right)^2 + 4V_{12}^2}. \end{aligned} \tag{A.1}$$

As in Chapter 1, the notation $\mathbf{x}^{(j)} \equiv (\mathbf{q}_j, \mathbf{Q}_1, \ldots, \mathbf{Q}_N)$ is used to indicate the position of ring polymer bead j and the full set of classical nuclei. The terms $V_{\mathrm{r}}(\mathbf{x}^{(j)})$ and $V_{\mathrm{p}}(\mathbf{x}^{(j)})$ are the molecular mechanics potential energy functions for the system with the hydride covalently bonded to the donor and acceptor atoms, respectively. The constant V_{12} =30.6 kcal/mol is fit to the experimental rate for the intrinsic reaction [3], and the product state potential $V_{\mathrm{p}}(\mathbf{x}^{(j)})$ includes a constant shift of $\Delta_{12} = +101.9$ kcal/mol to match the experimental driving force for the intrinsic reaction [3]. The difference in the value for Δ_{12} used here versus in Ref. [1] is due to different treatments of the atom-exclusions in the calculation of long-range electrostatic contributions.

Calculation of $V_{\mathrm{r}}(\mathbf{x}^{(j)})$ and $V_{\mathrm{p}}(\mathbf{x}^{(j)})$ is performed using a modified version of

the GROMOS 43A1 united atom forcefield [4]. These modifications, which again follow previous work [1, 5], are described in Figure A.2. A cutoff distance of 15 Å is applied to short-ranged non-bonding interactions, and electrostatic interactions beyond 9 Å are calculated using the particle mesh Ewald technique [6]. The bond-lengths for all non-transferring hydrogen atoms in the system are constrained using the SHAKE algorithm [7].

To avoid sampling configurations of the enzyme that are not relevant to the intrinsic reaction, weak harmonic restraints were introduced between the simulated enzyme and the reference crystal structure [8]. To avoid substrate dissociation during the long equilibrium sampling trajectories, weak harmonic restraints (k = 0.15 kcal mol^{-1}Å^{-2}) are applied to the heavy atoms in the glutamate moiety of the substrate and to the α-carbons of the neighboring α-helix segment composed of residues 26 to 35; to prevent large-scale conformation rearrangements in DHFR [9], weak harmonic restraints (k = 0.001 kcal mol^{-1}Å^{-2}) are applied to all other heavy atoms in the enzyme. Figure A.3 demonstrates that these restraints do not measurably affect the reactive trajectories used in our analysis of dynamical correlations.

A.2 Calculation Details

We initialize and equilibrate the system using the protocol described in Ref. [1]. The system is initialized from the DHFR crystal structure in the active configuration (PDB code: 1RX2) [8]. Crystallographic 2-mercaptoethanol and manganese ions are removed; crystallographic waters are not. The amine side chain of Q102 is rotated 180 degrees to correctly coordinate the adenine moiety of the cofactor [10]. To be consistent with the observed hydrogen bonding networks in the crystal structure, histidine residues 45, 124, and 149 are protonated at nitrogen ND1, histidine residues 114 and 141 are protonated at nitrogen NE2, and both DHFR cysteine residues are in their protonated form [1]. The enzyme is explicitly solvated using 4,122 SPC/E rigid water molecules [11] in a truncated octahedral

simulation cell with constant volume and periodic boundary conditions. The periodic image distance for the cell is 57.686 Å. Twelve Na^+ ions are included for charge neutrality. The full system includes N =14,080 classical nuclei.

All RPMD and classical MD trajectories are numerically integrated using the leap-frog integrator implemented in Gromacs-4.0.7. The simultaneous positions and velocities for each integration time-step in the trajectories are stored for analysis. Unless otherwise stated, the RPMD equations of motion are integrated using a time-step of 0.025 fs and classical MD trajectories are integrated using a time-step of 1 fs. The classical MD trajectories are used only for the initial equilibration of the system and for additional results presented here in the SI; all data presented in Chapter 1 are obtained using the quantized RPMD trajectories. Throughout the text, standard error estimates are calculated from five block-averages of the data.

From the initial geometry of the crystal structure, the system is equilibrated on the reactant potential energy surface V_r using classical MD. In a series of three equilibration steps, MD trajectories of length 10 ps in time are performed with progressively weaker harmonic restraints between the heavy atom positions and the crystal structure; the restraint force constants for the three equilibration runs are 100, 50, and 25 kcal $\mathrm{mol}^{-1}\mathrm{\AA}^{-2}$, respectively, and the runs are performed in the NVT ensemble using Berendsen thermostat with a coupling constant of 0.01 ps [12]. After initial equilibration to the reactants basin, the system is equilibrated on the full potential energy surface (Equation A.1) for an additional 100 ps of classical MD. Finally, the ring polymer representation for the quantized hydride was introduced at the geometry of the relaxed classical system and equilibrated for an additional 1 ps using RPMD with the Berendsen thermostat.

A.3 Free Energy Sampling

The free energy (FE) profile in Figure 1.1 is calculated as a function of the collective variable $\lambda(\mathbf{x}^\mathrm{c}) \equiv V_\mathrm{r}(\mathbf{x}^\mathrm{c}) - V_\mathrm{p}(\mathbf{x}^\mathrm{c})$, where $\mathbf{x}^\mathrm{c} \equiv (\mathbf{q}^\mathrm{c}, \mathbf{Q}_1, \ldots, \mathbf{Q}_N)$, and

$\mathbf{q}^{\mathrm{c}} = \sum_{j=1}^{n} \mathbf{q}_j/n$ is the ring polymer centroid of mass $m_{\mathrm{c}} = nm_n$. The umbrella sampling method [13] is used to efficiently sample this collective variable between the reactant and product basins. Independent RPMD sampling trajectories are performed using biasing potentials of the form

$$\sum_{j=1}^{n} \left[V_{\mathrm{EVB}}(\mathbf{x}^{(j)}) \right] + \frac{1}{2} k_l \left(\lambda \left(\mathbf{x}^{\mathrm{c}} \right) - \lambda_l \right)^2, \qquad l = 1, \ldots, 20, \tag{A.2}$$

where the $\{k_l\}$ and $\{\lambda_l\}$ are listed in Table A.1.

For the RPMD trajectories used to sample the FE profile, a ring polymer bead mass of m_n =$12/n$ amu was used to diminish the separation of timescales for the motion of the ring polymer and the rest of the system. Changing this parameter does not affect the ensemble of configurations that are sampled in the calculation of the FE profile; it merely allows for the sampling trajectories to be performed with a larger simulation time-step (0.1 fs) than is used in the dynamical trajectories. Furthermore, unlike the RPMD dynamical trajectories in which the long-range electrostatic contributions are updated every time-step, we use twin-ranged cut-offs [4] in the FE sampling trajectories such that non-bonding interactions beyond 9 Å are only updated every 5 fs. Sampling trajectories are performed at constant temperature using the velocity rescaling thermostat [14] with a relaxation time of 0.05 ps.

The sampling trajectories are initialized in order of increasing λ_l, as follows. The first sampling trajectory ($l = 1$) was initialized from the equilibrated system in the reactant basin. After 25 ps of simulation, the configuration from this first sampling trajectory was used to initialize the second sampling trajectory ($l = 2$). After 25 ps of simulation, the configuration from the second trajectory was used to initialize the third sampling trajectory ($l = 3$), and so on. A total simulation length of 6 ns is sampled for each value of l, and the weighted histogram analysis method (WHAM) [15] is used to calculate the unbiased FE profile $F(\lambda)$ from the set of sampling trajectories.

To improve the overlap of the trajectories in coordinates other than $\lambda\left(\mathbf{x}^{\mathrm{c}}\right)$, we

follow the swapping procedure described by Warshel and coworkers [16]. Configurations are swapped between neighboring values of windows every 100 ps, and the first 25 ps after a swap are discarded. Comparing results obtained with and without the use of this swapping protocol, we found no significant difference in the calculated free energy profile (Figure A.4). Each sampling trajectory for the calculation of the FE profile without swapping was also of length 6 ns in time.

In addition to calculating the quantized FE profile using the RPMD sampling trajectories, we repeat the sampling protocol with classical MD trajectories to obtain the classical FE profile for the intrinsic reaction. Figure A.5A compares these two results; the results for the quantized system are identical to those from Figure 1.1B.

For the calculation of equilibrium ensemble averages, the configurations of the enzyme are aligned to remove overall translational and orientational diffusion. As in previous studies [1], this is done using the following iterative protocol. In a first step, all configurations in the ensemble are aligned to the crystal structure, and the atom positions of the aligned structure are averaged. In a second step, all of the configurations in the ensemble are aligned to the average structure from the first step. In all cases, the RMSD between the average structures of subsequent iterations converged to within 10^{-7} Å in less than 20 iterations.

A.4 The Dividing Surface Ensemble

Boltzmann-weighted sampling on the reaction dividing surface ($\lambda(\mathbf{x}^{c}) = $ -4.8 kcal/mol) is performed with constrained molecular dynamics using the SHAKE algorithm [7]. The existing implementation of SHAKE in Gromacs-4.0.7 is modified to constrain both classical MD and RPMD with respect to the collective variable $\lambda(\mathbf{x}^{c})$. To remove the hard-constraint bias from the ensemble of configurations that is sampled in the constrained dynamics [17, 18], each sampled configuration is weighted

by $[H(\mathbf{x}^{c})]^{-1/2}$, where

$$H(\mathbf{x}^{c}) = (m_{c})^{-1} \left|\nabla_{\mathbf{q}^{c}} \lambda(\mathbf{x}^{c})\right|^{2} + \sum_{k=1}^{N} M_{k}^{-1} \left|\nabla_{\mathbf{Q}_{k}} \lambda(\mathbf{x}^{c})\right|^{2}. \tag{A.3}$$

Seven long, independent RPMD trajectories are run with the dividing surface constraint. These constrained trajectories are initialized from configurations obtained in the umbrella sampling trajectories that are restrained to the dividing surface region using Equation A.2, and they are performed at constant temperature using the velocity rescaling thermostat [14] with a relaxation time of 0.05 ps. Following an initial equilibration of 25 ps, each of the constrained trajectories is run for 1 ns, and dividing surface configurations are sampled every 4 ps. As with the umbrella sampling trajectories, the constrained dynamics utilize a ring-polymer bead mass of m_n =$12/n$ amu to enable a time-step of 0.1 fs.

To eliminate the effects of overall translational and rotational diffusion from the analysis of the reactive trajectories, the phase-space points for the reactive trajectories are aligned at $t = 0$ (i.e., the point of initial release from the dividing surface). This is done exactly as in the calculation of equilibrium averages. Using the ensemble of configurations corresponding to reactive trajectories at $t = 0$, the rotation and translation for each particular trajectory is determined. This translation and rotation is applied to the configuration of each time-step in the trajectory, and only the rotation is applied to the velocities at each time-step in the trajectory.

A.5 Calculation of the Transmission Coefficient

Using a dividing surface of $\lambda(\mathbf{x}^{c}) = \lambda^{\ddagger}$, the time-dependent transmission coefficient for the reaction is [19, 20, 21, 22]

$$\kappa(t) = \frac{\left\langle \dot{\lambda}(\mathbf{x}^{c}(0), \dot{\mathbf{x}}^{c}(0))\, \theta\big(\lambda(\mathbf{x}^{c}(t)) - \lambda^{\ddagger}\big) \right\rangle_{\lambda^{\ddagger}}}{\left\langle \dot{\lambda}(\mathbf{x}^{c}(0), \dot{\mathbf{x}}^{c}(0))\, \theta\Big(\dot{\lambda}(\mathbf{x}^{c}(0), \dot{\mathbf{x}}^{c}(0))\Big) \right\rangle_{\lambda^{\ddagger}}} \tag{A.4}$$

where $\langle \ldots \rangle_{\lambda^\ddagger}$ denotes the Boltzmann-weighted distribution of configurations on the dividing surface, and $\theta(x)$ is the Heaviside function. The transmission coefficient is evaluated by initializing RPMD trajectories from configurations ($\mathbf{x}^c(0)$) on the dividing surface with velocities ($\dot{\mathbf{x}}^c(0)$) drawn from the Maxwell-Boltzmann distribution. These initial configurations are then correlated with the configurations ($\mathbf{x}^c(t)$) reached by the unconstrained RPMD trajectories after evolving the dynamics in the NVE ensemble for time t.

In the current study, each sampled configuration on the dividing surface is used to generate three unconstrained RPMD trajectories that are evolved both forward and backwards in time for 100 fs, such that 10,500 half-trajectories are released from the dividing surface. The initial velocities for each trajectory are drawn independently; these time-zero velocities, the time-zero positions, and the corresponding time-zero forces at the initial positions are used to initialize the leap-frog integrator.

In addition to calculating the transmission coefficient for the quantized hydride transfer using RPMD, we repeat the protocol with classical MD trajectories to obtain the classical transmission coefficient. Figure A.5B compares these two results; the results for the quantized system are identical to those from Figure 1.1C.

A.6 Additional Measures of Dynamical Correlations

To confirm that Figure 1.2, which considers only heavy atom positions, did not neglect important dynamical correlations, we include the corresponding plots with all atoms for the enzyme included (Figure A.6). To examine the robustness of our conclusions about dynamical correlations in the system, we present various alternative measures of dynamical correlations in Figures A.7, A.8, and A.9. Figure A.7 presents alternative methods for analyzing the dynamical correlation measure $d_{ij}(t)$. In Figure A.8, we present a measure of dynamical correlations that are

non-local in time,

$$d_{ij}^{\Delta t}(t) = \frac{\langle \mathbf{v}_i(t) \cdot \mathbf{v}_j(t+\Delta t) \rangle_t}{\left(\langle |\mathbf{v}_i(t)|^2 \rangle_t \langle |\mathbf{v}_j(t+\Delta t)|^2 \rangle_t \right)^{1/2}} - \frac{\langle \mathbf{v}_i(t) \cdot \mathbf{v}_j(t+\Delta t) \rangle}{\left(\langle |\mathbf{v}_i|^2 \rangle \langle |\mathbf{v}_j|^2 \rangle \right)^{1/2}}, \tag{A.5}$$

where $\langle ... \rangle_t$ and $\langle ... \rangle$ are defined as in Chapter 1. In Figure A.9, we present a measure of dynamical correlations between perpendicular components of the velocity,

$$d_{ij}^{\perp,\xi_1,\xi_2}(t) = \frac{D_{ij}^{\perp,\xi_1,\xi_2}(t)}{\left(D_{ii}^{\perp,\xi_1,\xi_1}(t) D_{jj}^{\perp,\xi_2,\xi_2}(t) \right)^{1/2}}, \tag{A.6}$$

where

$$D_{ij}^{\perp,\xi_1,\xi_2}(t) = \left\langle \left(\bar{v}_i^{\xi_1}(t) - \left\langle \bar{v}_i^{\xi_1} \right\rangle \right) \left(\bar{v}_j^{\xi_2}(t) - \left\langle \bar{v}_j^{\xi_2} \right\rangle \right) \right\rangle_t \tag{A.7}$$

and where $\bar{\mathbf{v}}_i = (\bar{v}_i^1, \bar{v}_i^2, \bar{v}_i^3)$ is the absolute velocity vector in Cartesian coordinates. As for the measures presented in Chapter 1, dynamical correlations are found to be localized in the substrate and cofactor regions, with only weak signatures in the protein residues surrounding the active site.

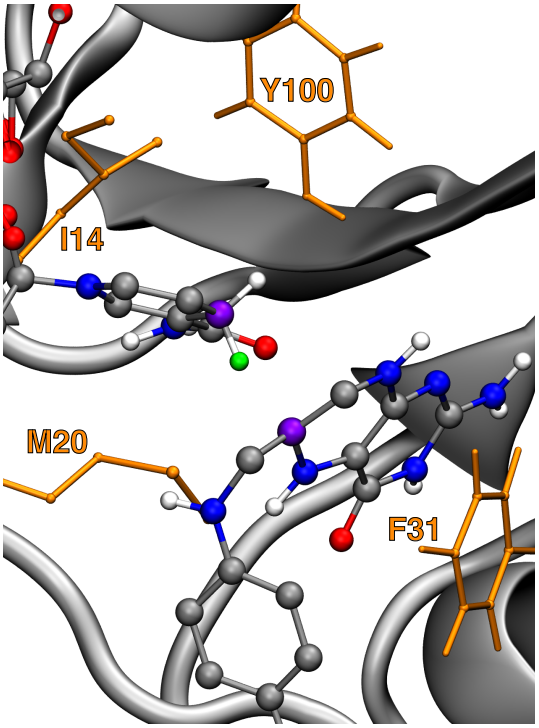

Figure A.1: The active site region of the DHFR enzyme, with the transferring hydride (green) in the reactant state, the donor and acceptor carbon atoms in purple, and relevant protein residues in gold.

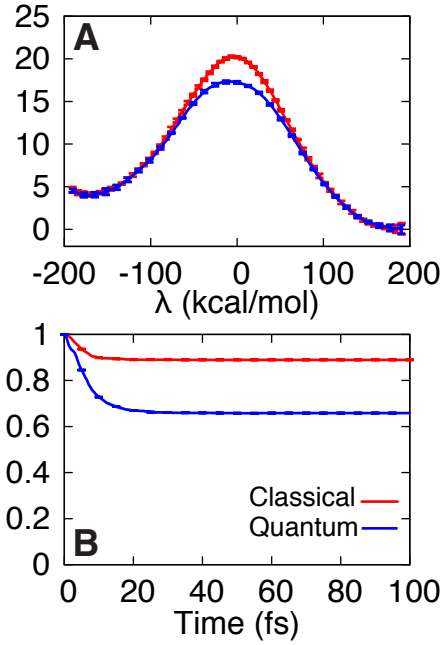

Figure A.2: (A) The quantized and classical free energy profiles for the reaction. (B) The quantized and classical time-dependent transmission coefficient corresponding to the dividing surface at $\lambda(\mathbf{x}) = -4.8$ kcal/mol.

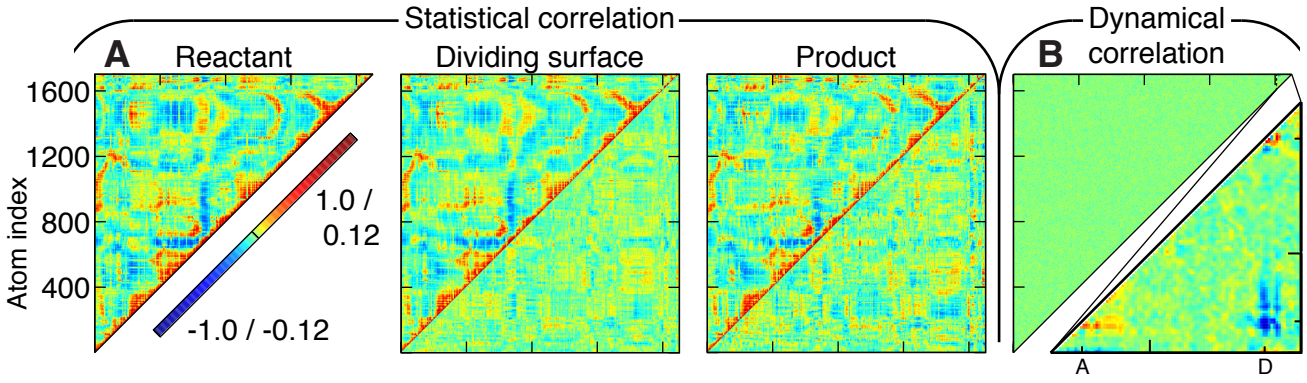

Figure A.3: (A) (upper triangles) The covariance c_{ij} among position fluctuations in DHFR, plotted for the reactant, dividing surface, and product regions. All atoms are indexed according to PDB 1RX2. (lower triangles) The difference with respect to the plot for the reactant basin. (B) (upper triangle) The integrated dynamical correlation measure d_{ij}, indexed as in (A). (lower triangle) The substrate and cofactor regions are enlarged. Dynamical correlations between atom-pairs that share a holonomic constraint are excluded from part B. Comparison of the current figure (which includes all atoms) with Figure 1.2 (which includes only heavy atoms) leaves the conclusions from Chapter 1 unchanged.

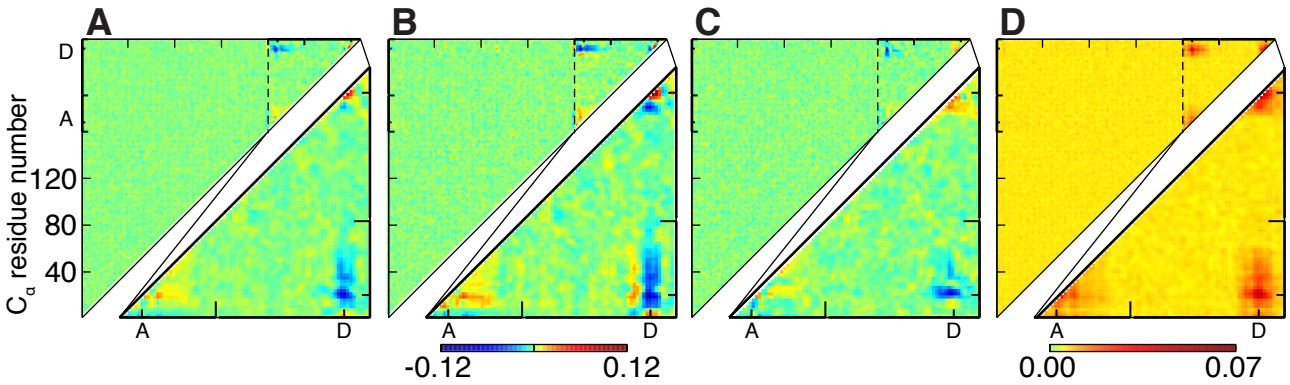

Figure A.4: Alternative measures of the dynamical correlation. (A) The integrated dynamical correlation measure $d_{ij} = \int_{-\tau}^{\tau} dt\, d_{ij}(t)$, reproduced from Figure 1.2E. (B) Including only the forward-integrated time, $\int_{0}^{\tau} dt\, d_{ij}(t)$. ($C$) Including only the backward-integrated time, $d_{ij} = \int_{-\tau}^{0} dt\, d_{ij}(t)$. ($D$) Including the integrated absolute value, $\int_{-\tau}^{\tau} dt\ |d_{ij}(t)|$. In all cases, $\tau = 100$ fs.

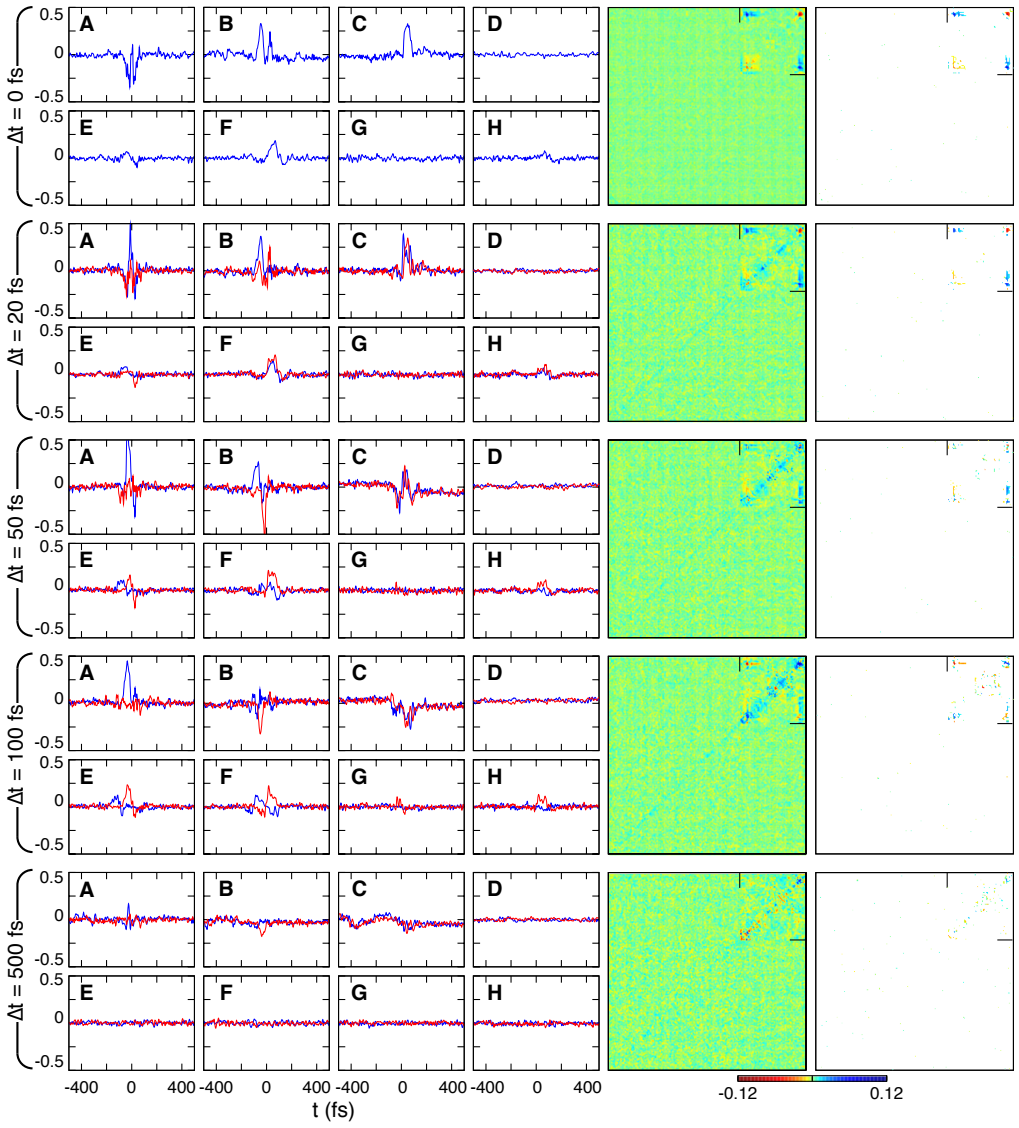

Figure A.5: (*A*-*H*) The measure of temporally non-local dynamical correlations, $d_{ij}^{\Delta t}(t)$, plotted at various separation times Δt and for various atom pairs, including (A) the donor and acceptor atom pair, (B) the substrate-based C_7 and the acceptor atom pair, (C) the cofactor-based C_{N4} and donor atom pair, (D) the active site-based protein atoms that define the distance Θ_2 in Figure 1.4B, (E) the active site based backbone I4 O and the acceptor atom pair, (F) the active site based backbone I4 O and the donor atom pair, (G) the active site based backbone G95 C_α and the acceptor atom pair, and (H) the active site based backbone G95 C_α and the donor atom pair. Two curves (red and blue) are plotted, since $d_{ij}^{\Delta t}(t)$ is not symmetric with respect to atom indices i and j for non-zero Δt. At right, the integrated measure $d_{ij}^{\Delta t} = \int_{-\tau}^{\tau} dt\, d_{ij}^{\Delta t}(t)$ for each lag time, plotted as a function of the protein α-carbon atoms and the heavy atoms of the substrate and cofactor, as in Figure 1.2E. In all cases, $\tau = 100$ fs. At far right, the same integrated measure is replotted, only displaying data points for which the magnitude of the integrated measure exceeds twice the estimated standard error.

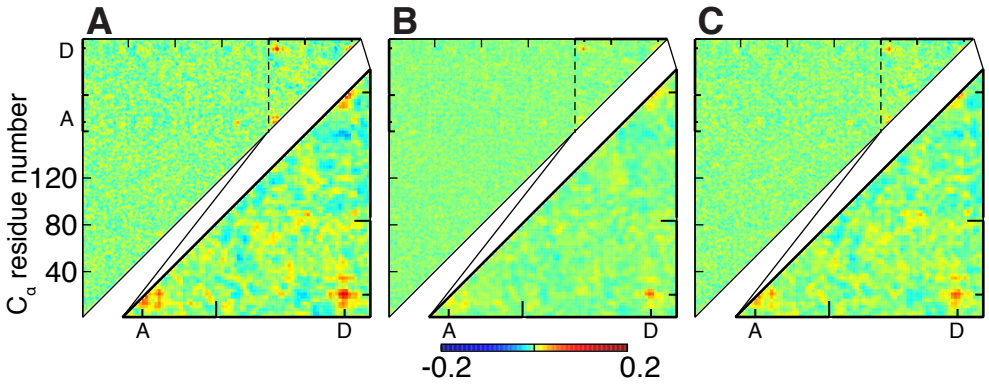

Figure A.6: The measure of dynamical correlations between perpendicular components of the velocity, $d_{ij}^{\perp,\xi_1,\xi_2}(t)$. (A) The integrated measure $d_{ij}^{\perp} = \sum_{\xi_1,\xi_2=1}^{3} \int_{-\tau}^{\tau} dt\, d_{ij}^{\perp,\xi_1,\xi_2}(t)$ that includes all components. (B) The integrated measure $d_{ij}^{\perp} = \sum_{\xi}^{3} \int_{-\tau}^{\tau} dt\, d_{ij}^{\perp,\xi,\xi}(t)$ that includes only diagonal components. (C) The integrated measure $d_{ij}^{\perp} = \sum_{\xi_1\neq\xi_2=1}^{3} \int_{-\tau}^{\tau} dt\, d_{ij}^{\perp,\xi_1,\xi_2}(t)$ that includes only off-diagonal components. In all cases, $\tau = 100$ fs, and the integrated measures are plotted as a function of the protein α-carbon atoms and the heavy atoms of the substrate and cofactor, as in Figure 1.2E.

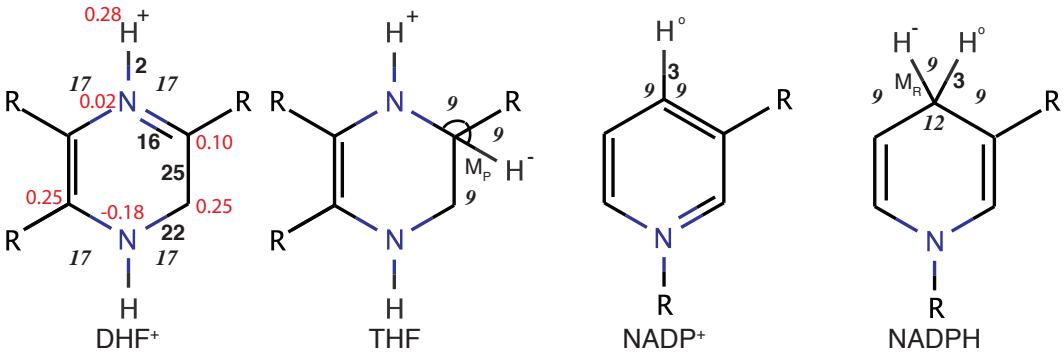

Figure A.7: Modifications to the GROMOS 43A1 united atom forcefield only involve the DHF^+, THF, $NADP^+$, and NADPH species. The resulting potential energy surface is as close as possible to that used in earlier studies of statistical correlation in DHFR hydride transfer catalysis [1]. Only parameters that differ from the GROMOS forcefield are indicated; parameters shown for DHF^+ differ with respect to those for DHF. In red, the atomic charges for DHF^+ are shown. In black, the bond-type (bold) and angle-type (italics) indices for the GROMOS forcefield are shown. Explicit representations are used for the transferring hydride (H^-) in the THF and NADPH species, the pro-*S* hydrogen atom (H^o) in the NADPH and $NADP^+$ species, and the proton (H^+) attached to the nearest neighbor of the donor carbon. Firstly, we describe the treatment of H^-. The transferring hydride interacts with the donor and acceptor carbons via Morse potentials M_R and M_P, respectively [1]. Following GROMOS convention, non-bonding interactions are excluded between H^- and its 1^{st}-, 2^{nd}- and 3^{rd}-nearest neighbors, defined in terms of bonding connectivity. Additionally, non-bonding interactions between H^- and the H^o, donor, and acceptor atoms are excluded, regardless of the local bonding environment of the H^- atom. Secondly, we describe the treatment of H^o. The bond length for H^o is constrained to a fixed value of 1.09 Å, and planarity of H^o with respect to the nicotinamide ring in $NADP^+$ is enforced using the planar improper dihedral angle potential in GROMOS. As for the hydride, non-bonding interactions are excluded between H^o and its 1^{st}-, 2^{nd}- and 3^{rd}-nearest neighbors. Thirdly, we describe the treatment of H^+. Nonbonding interactions are excluded between H^+ and its 1^{st}- and 2^{nd}- nearest neighbors; 3^{rd}-nearest neighbor nonbonding interactions are treated through using a 1-4 potential. For the LJ interactions involving these explicit hydrogen atoms, H^+ is treated using the parameters for a charged hydrogen, and both H^o and H^- are treated using the parameters for a hydrogen bound to a carbon. For the LJ interactions involving the donor and acceptor, these atoms are treated using the parameters for a bare carbon atom.

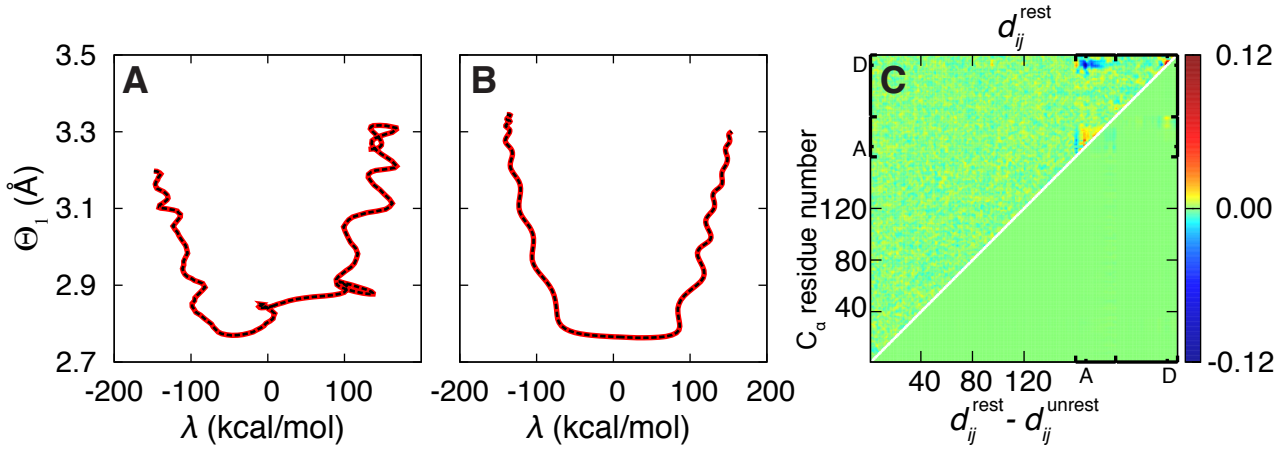

Figure A.8: Tests of the degree to which the weak harmonic restraints impact the dynamics of the reactive trajectories. (A) Comparison of two trajectories that are initialized from the same positions and velocities on the dividing surface, which are calculated with (red) and without (black) the weak restraints. Trajectories are evolved for a total of 200 fs and are viewed in the plane of the donor-acceptor distance ($\Theta_1(\mathbf{x}^c)$) and the reaction progress variable ($\lambda(\mathbf{x}^c)$). (B) In red, the mean pathway from the ensemble of 750 reactive trajectories, $\sigma = (\langle\lambda(\mathbf{x}^c)\rangle_t, \langle\Theta_1(\mathbf{x}^c)\rangle_t)$, calculated for trajectories obtained with the weak restraints. In black, the mean pathway from the ensemble of 750 trajectories that are initialized from the same phase-space points but which do not include the weak restraints. (C) (pper triangle) The dynamical correlation measure d_{ij}^{rest}, calculated from an ensemble of 750 reactive trajectories using the weak harmonic restraints; this quantity is identical to the that reported in Figure 1.2E, except that fewer trajectories are used here. (lower triangle) The difference between d_{ij}^{rest} and d_{ij}^{unrest}, which is calculated from the ensemble of 750 trajectories that are initialized from the same phase-space points but which do not include the weak restraints. For all three tests, the impact of the weak restraints on the dynamics of the reactive trajectories is undetectable.

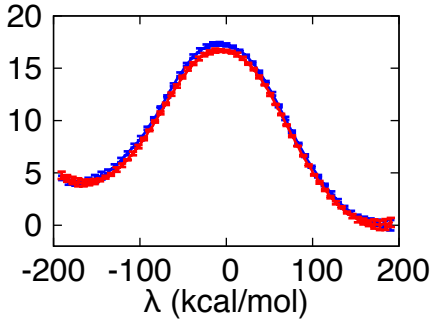

Figure A.9: Free energy profiles obtained with (red) and without (blue) swapping of configurations from neighboring 6 ns sampling trajectories in the WHAM calculation.

l	λ_l	k_l
1	-188.7	0.002
2	-154.0	0.002
3	-120.0	0.002
4	-87.1	0.002
5	-57.4	0.004
6	-56.1	0.002
7	-35.8	0.004
8	-18.3	0.004
9	-17.8	0.008
10	-6.4	0.008
11	1.4	0.004
12	1.5	0.008
13	4.3	0.006
14	14.2	0.004
15	23.3	0.002
16	34.2	0.004
17	57.9	0.002
18	95.8	0.002
19	135.7	0.002
20	170.0	0.000

Table A.1: The umbrella sampling parameters $\{\lambda_l\}$ and $\{k_l\}$ in units of kcal/mol and mol/kcal, respectively.

References

[1] P. K. Agarwal, S. R. Billeter, and S. Hammes-Schiffer. Nuclear quantum effects and enzyme dynamics in dihydrofolate reductase catalysis. *J. Phys. Chem. B*, 106(12):3283–3293, 2002.

[2] A. Warshel and R. M. Weiss. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Am. Chem. Soc.*, 102(20):6218–6226, 1980.

[3] C. A. Fierke, K. A. Johnson, and S. J. Benkovic. Construction and evaluation of the kinetic scheme associated with dihydrofolate-reductase from *Escherichia coli*. *Biochemistry*, 26(13):4085–4092, 1987.

[4] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger Mark, A. E., W. R. P. Scott, and I. G. Tironi. *Biomolecular Simulation: The GROMOS96 manual and user guide*. Hochschuleverlag AG an der ETH Zürich, Zürich, Switzerland, 1996.

[5] K. F. Wong, J. B. Watney, and S. Hammes-Schiffer. Analysis of electrostatics and correlated motions for hydride transfer in dihydrofolate reductase. *J. Phys. Chem. B*, 108(32):12231–12241, 2004.

[6] T. Darden, D. York, and L. Pedersen. Particle mesh ewald—An N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[7] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical-integration

of cartesian equations of motion of a system with constraints—Molecular-dynamics of N-alkanes. *J. Comput. Phys.*, 23(3):327–341, 1977.

[8] M. R. Sawaya and J. Kraut. Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: Crystallographic evidence. *Biochemistry*, 36(3):586–603, 1997.

[9] K. Arora and C. L. Brooks. Functionally important conformations of the met20 loop in dihydrofolate reductase are populated by rapid thermal fluctuations. *J. Am. Chem. Soc.*, 131(15):5642–5647, 2009.

[10] I. V. Khavrutskii, D. J. Price, J. Lee, and C. L. Brooks. Conformational change of the methionine 20 loop of *Escherichia coli* dihydrofolate reductase modulates pk(a) of the bound dihydrofolate. *Protein Sci.*, 16(6):1087–1100, 2007.

[11] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987.

[12] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and J. R. Haak. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[13] G. M. Torrie and J. P. Valleau. Non-physical sampling distributions in Monte-Carlo free-energy estimation—Umbrella sampling. *J. Comput. Phys.*, 23(2):187–199, 1977.

[14] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101, 2007.

[15] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.*, 13(8):1011–1021, 1992.

[16] H. B. Liu and A. Warshel. Origin of the temperature dependence of isotope effects in enzymatic reactions: The case of dihydrofolate reductase. *J. Phys. Chem. B*, 111(27):7852–7861, 2007.

[17] D. A. Zichi, G. Ciccotti, J. T. Hynes, and M. Ferrario. Molecular-dynamics simulation of electron-transfer reactions in solution. *J. Phys. Chem.*, 93(17):6261–6265, 1989.

[18] D. Frenkel and B. Smit. *Understanding molecular simulation: From algorithms to applications*. Academic Press, San Diego, 2nd edition, 2002.

[19] H. Bennett Charles. Molecular dynamics and transition state theory: The simulation of infrequent events. In *Algorithms for Chemical Computations*, volume 46 of *ACS Symposium Series*, pages 63–97. American Chemical Society, 1977. doi:10.1021/bk-1977-0046.ch004.

[20] D. Chandler. Statistical-mechanics of isomerization dynamics in liquids and transition-state approximation. *J. Chem. Phys.*, 68(6):2959–2970, 1978.

[21] I. R. Craig and D. E. Manolopoulos. A refined ring polymer molecular dynamics theory of chemical reaction rates. *J. Chem. Phys.*, 123(3):034102, 2005.

[22] I. R. Craig and D. E. Manolopoulos. Chemical reaction rates from ring polymer molecular dynamics. *J. Chem. Phys.*, 122(8):084106, 2005.

Appendix B

Appendix 2: Supplemental Information for Chapter 2

B.1 Potential Energy Surface

The potential energy surface used in simulations of kinetic isotope effects is the same as that described above (Equation A.7) with a parameterization term to correct for overestimation of the kinetic isotope effect (At 280 K, $k_H/k_T = 40\pm3$, compared to an experimental value of 6.1) . To correct this overestimation, the curvature of the Morse potentials used in the study in Chapter A.7 were scaled by a parameter γ such that

$$V_{\mathrm{M}}\left(r_{\mathrm{H,D/A}}, \gamma\alpha, De, r_0\right) = De\left(1 - e^{-\gamma\alpha(r-r_0)}\right)^2, \tag{B.1}$$

where $r_{\mathrm{H,D/A}}$ is distance between the hydride and the donor or acceptor carbon, $\gamma\alpha$ is scaled decay constant for the Morse potential, De is the dissociation energy of the bond, and r_0 is the equilibrium bond length. Here, for the donor carbon, $\alpha = 1.785$ A^{-1}, $De = 103$ kcal/mol, and $r_0 = 1.09$ A; for the acceptor carbon, $\alpha = 1.758$ A^{-1}, $De = 103$ kcal/mol, and $r_0 = 1.09$ A. γ has a value of 0.5 for both donor and acceptor carbons. The potential energy function was reparameterized to reproduce the H/T kinetic isotope effect at 280 K. The same potential energy function was used for each mass and temperature simulation.

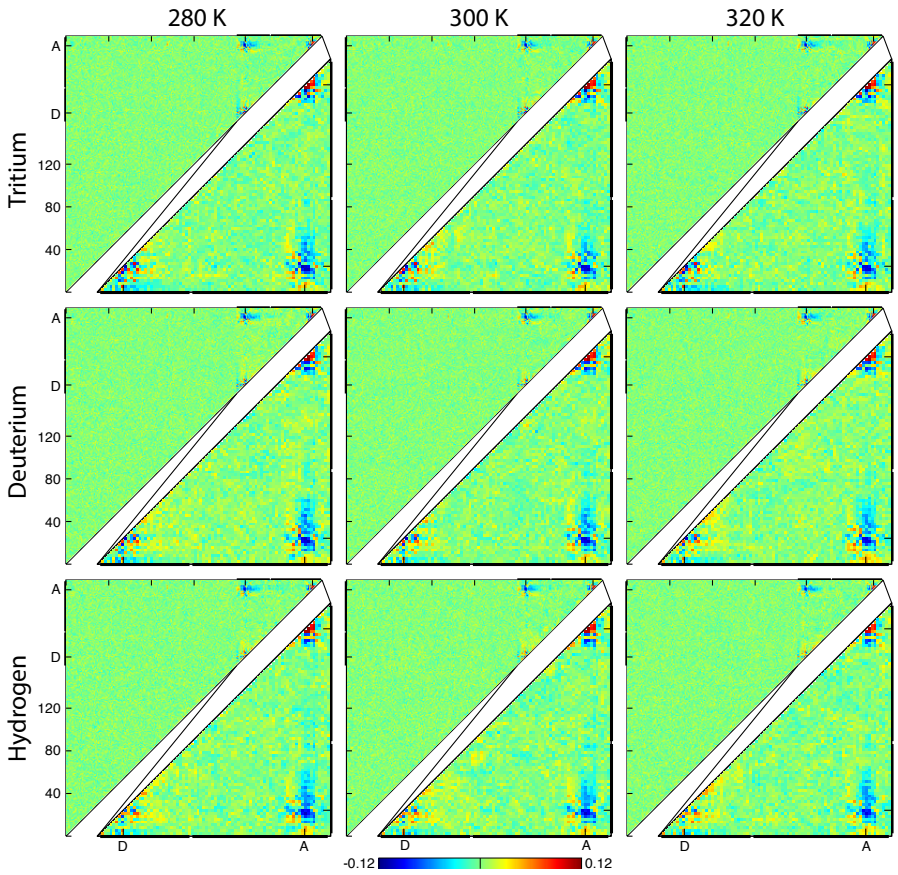

Figure B.1: The d_{ij} matrix is presented as in Figure 2.5 in the main text for each of the simulations.

l	λ_l	k_l
1	-141.245	0.0015
2	-116.972	0.0017
3	-93.900	0.0020
4	-72.250	0.0025
5	-53.342	0.0031
6	-36.848	0.0039
7	-25.951	0.0048
8	-18.137	0.0058
9	-12.316	0.0070
10	-8.016	0.0083
11	-4.841	0.0100
12	-2.631	0.0122
13	-1.684	0.0130
14	-0.925	0.0132
15	0.042	0.0105
16	2.259	0.0086
17	5.900	0.0071
18	11.066	0.0058
19	18.477	0.0046
20	29.542	0.0036
21	47.697	0.0027
22	69.746	0.0020
23	96.196	0.0016
24	125.000	0.0013
25	154.964	0.0013

Table B.1: The umbrella sampling parameters $\{\lambda_l\}$ and $\{k_l\}$ in units of kcal/mol and mol/kcal, respectively.

Appendix C

Appendix 3: Supplemental Information for Chapter 3

Table C.1: Semi-empirical parameters from AM1 and reparameterized values AM1-SRP for the eight atom types defined here.

Hydrogen			
Parameter (Units)	AM1	AM1-SPR	Δ
U_{ss} (eV)	-11.3964270	-10.9728011	-0.4236259
β_s (eV)	-6.1737870	-6.2208155	0.0470285
ζ_s (eV)	1.1880780	1.1880780	0.0000000
α (Å$^-$1)	2.8823240	2.8770609	0.0052631
G_{ss} (eV)	12.8480000	13.4647988	-0.6167988
K_1 (eV)	0.1227960	0.1376800	-0.0148840
L_1 (1/Å^2)	5.0000000	5.0606517	-0.0606517
M_1 (Å)	1.2000000	1.1597789	0.0402211
K_2 (eV)	0.0050900	0.0043265	0.0007635
L_2 (1/Å^2)	5.0000000	4.9968359	0.0031641
M_2 (Å)	1.8000000	1.7820358	0.0179642
K_3 (eV)	-0.0183360	-0.0191089	0.0007729
L_3 (1/Å^2)	2.0000000	2.0316003	-0.0316003
M_3 (Å)	2.1000000	1.8595325	0.2404675
Carbon			

Continued on next page

Table C.1 – *Continued from previous page*

Parameter (Units)	AM1	AM1-SPR	Δ
Parameter (Units)	AM1	AM1-SPR	Δ
U_{ss} (eV)	-52.0286580	-48.7063335	-3.3223245
U_{pp} (eV)	-39.6142390	-41.4299458	1.8157068
β_s (eV)	-15.7157830	-16.0743327	0.3585497
β_p (eV)	-7.7192830	-7.8428796	0.1235966
ζ_s (eV)	1.8086650	1.8086650	0.0000000
ζ_p (eV)	1.6851160	1.6851160	0.0000000
α (Å$^-$1)	2.6482740	2.6827466	-0.0344726
G_{ss} (eV)	12.2300000	10.7562544	1.4737456
G_{sp} (eV)	11.4700000	11.9872691	-0.5172691
G_{pp} (eV)	11.0800000	11.8931729	-0.8131729
$G_{pp'}$ (eV)	9.8400000	10.7471996	-0.9071996
H_{sp} (eV)	2.4300000	2.4514667	-0.0214667
K_1 (eV)	0.0113550	0.0099602	0.0013948
L_1 (1/Å^2)	5.0000000	4.9981346	0.0018654
M_1 (Å)	1.6000000	1.5852376	0.0147624
K_2 (eV)	0.0459240	0.0396675	0.0062565
L_2 (1/Å^2)	5.0000000	5.0009459	-0.0009459
M_2 (Å)	1.8500000	1.8541506	-0.0041506
K_3 (eV)	-0.0200610	-0.0225580	0.0024970
L_3 (1/Å^2)	5.0000000	5.0107397	-0.0107397
M_3 (Å)	2.0500000	1.9965302	0.0534698
K_4 (eV)	-0.0012600	-0.0010710	-0.0001890
L_4 (1/Å^2)	5.0000000	4.9966202	0.0033798
M_4 (Å)	2.6500000	2.6125005	0.0374995
Nitrogen			
Parameter (Units)	AM1	AM1-SPR	Δ
U_{ss} (eV)	-71.8600000	-71.9816390	0.1216390
U_{pp} (eV)	-57.1675810	-57.6291203	0.4615393
β_s (eV)	-20.2991100	-20.4073334	0.1082234
β_p (eV)	-18.2386660	-18.0860754	-0.1525906
ζ_s (eV)	2.3154100	2.3154100	0.0000000
ζ_p (eV)	2.1579400	2.1579400	0.0000000
α (Å$^-$1)	2.9472860	2.5182816	0.4290044

Continued on next page

Table C.1 – *Continued from previous page*

Parameter (Units)	AM1	AM1-SPR	Δ
G_{ss} (eV)	13.5900000	13.6468825	-0.0568825
G_{sp} (eV)	12.6600000	12.4115008	0.2484992
G_{pp} (eV)	12.9800000	12.8976040	0.0823960
$G_{pp'}$ (eV)	11.5900000	11.3749638	0.2150362
H_{sp} (eV)	3.1400000	3.1449539	-0.0049539
K_1 (eV)	0.0252510	0.0260333	-0.0007823
L_1 (1/Å^2)	5.0000000	4.9967258	0.0032742
M_1 (Å)	1.5000000	1.5403256	-0.0403256
K_2 (eV)	0.0289530	0.0332843	-0.0043313
L_2 (1/Å^2)	5.0000000	4.9912738	0.0087262
M_2 (Å)	2.1000000	2.1030025	-0.0030025
K_3 (eV)	-0.0058060	-0.0049351	-0.0008709
L_3 (1/Å^2)	2.0000000	2.0223028	-0.0223028
M_3 (Å)	2.4000000	2.2711710	0.1288290
Oxygen			
Parameter (Units)	AM1	AM1-SPR	Δ
U_{ss} (eV)	-97.8300000	-97.5085847	-0.3214153
U_{pp} (eV)	-78.2623800	-80.2223761	1.9599961
β_s (eV)	-29.2727730	-29.5107658	0.2379928
β_p (eV)	-29.2727730	-29.4773322	0.2045592
ζ_s (eV)	3.1080320	3.1080320	0.0000000
ζ_p (eV)	2.5240390	2.5240390	0.0000000
α (Å$^-$1)	4.4553710	4.7567420	-0.3013710
G_{ss} (eV)	15.4200000	14.0270027	1.3929973
G_{sp} (eV)	14.4800000	13.9863945	0.4936055
G_{pp} (eV)	14.5200000	15.3378125	-0.8178125
$G_{pp'}$ (eV)	12.9800000	13.2040980	-0.2240980
H_{sp} (eV)	3.9400000	4.0224664	-0.0824664
K_1 (eV)	0.2809620	0.2388177	0.0421443
L_1 (1/Å^2)	5.0000000	5.0085567	-0.0085567
M_1 (Å)	0.8479180	0.7207303	0.1271877
K_2 (eV)	0.0814300	0.0809051	0.0005249
L_2 (1/Å^2)	7.0000000	7.0047673	-0.0047673
M_2 (Å)	1.4450710	1.5517439	-0.1066729

Continued on next page

Table C.1 – *Continued from previous page*

Parameter (Units)	AM1	AM1-SPR	Δ
NADPH-C2			
Parameter (Units)	AM1	AM1-SPR	Δ
U_{ss} (eV)	-52.0286580	-58.5735761	6.5449181
U_{pp} (eV)	-39.6142390	-39.0487175	-0.5655215
β_s (eV)	-15.7157830	-16.4489775	0.7331945
β_p (eV)	-7.7192830	-9.2631396	1.5438566
ζ_s (eV)	1.8086650	1.8086650	0.0000000
ζ_p (eV)	1.6851160	1.6851160	0.0000000
α (Å$^-$1)	2.6482740	2.8222911	-0.1740171
G_{ss} (eV)	12.2300000	12.2550814	-0.0250814
G_{sp} (eV)	11.4700000	11.4903395	-0.0203395
G_{pp} (eV)	11.0800000	11.1693256	-0.0893256
$G_{pp'}$ (eV)	9.8400000	9.6962720	0.1437280
H_{sp} (eV)	2.4300000	2.4435628	-0.0135628
K_1 (eV)	0.0113550	0.0141867	-0.0028317
L_1 (1/Å^2)	5.0000000	5.0045311	-0.0045311
M_1 (Å)	1.6000000	1.5394964	0.0605036
K_2 (eV)	0.0459240	0.0572678	-0.0113438
L_2 (1/Å^2)	5.0000000	5.0235534	-0.0235534
M_2 (Å)	1.8500000	1.5368198	0.3131802
K_3 (eV)	-0.0200610	-0.0250762	0.0050152
L_3 (1/Å^2)	5.0000000	4.9807597	0.0192403
M_3 (Å)	2.0500000	2.4772702	-0.4272702
K_4 (eV)	-0.0012600	-0.0015750	0.0003150
L_4 (1/Å^2)	5.0000000	5.0039299	-0.0039299
M_4 (Å)	2.6500000	2.5936848	0.0563152
PYR-C2			
Parameter (Units)	AM1	AM1-SPR	Δ
U_{ss} (eV)	-52.0286580	-56.0388083	4.0101503
U_{pp} (eV)	-39.6142390	-39.8722743	0.2580353
β_s (eV)	-15.7157830	-18.8116388	3.0958558
β_p (eV)	-7.7192830	-7.4432109	-0.2760721
ζ_s (eV)	1.8086650	1.8086650	0.0000000
ζ_p (eV)	1.6851160	1.6851160	0.0000000

Continued on next page

Table C.1 – *Continued from previous page*

Parameter (Units)	AM1	AM1-SPR	Δ
α (Å$^-$1)	2.6482740	3.1779288	-0.5296548
G_{ss} (eV)	12.2300000	12.1294519	0.1005481
G_{sp} (eV)	11.4700000	11.7926331	-0.3226331
G_{pp} (eV)	11.0800000	11.9203871	-0.8403871
$G_{pp'}$ (eV)	9.8400000	9.3157590	0.5242410
H_{sp} (eV)	2.4300000	2.4242747	0.0057253
K_1 (eV)	0.0113550	0.0141927	-0.0028377
L_1 (1/Å^2)	5.0000000	5.0103654	-0.0103654
M_1 (Å)	1.6000000	1.6339995	-0.0339995
K_2 (eV)	0.0459240	0.0574002	-0.0114762
L_2 (1/Å^2)	5.0000000	5.0174729	-0.0174729
M_2 (Å)	1.8500000	1.6575542	0.1924458
K_3 (eV)	-0.0200610	-0.0185708	-0.0014902
L_3 (1/Å^2)	5.0000000	5.0032824	-0.0032824
M_3 (Å)	2.0500000	2.3024310	-0.2524310
K_4 (eV)	-0.0012600	-0.0009520	-0.0003080
L_4 (1/Å^2)	5.0000000	5.0009864	-0.0009864
M_4 (Å)	2.6500000	2.6114740	0.0385260
H193-NE2			
Parameter (Units)	AM1	AM1-SPR	Δ
U_{ss} (eV)	-71.8600000	-73.6506122	1.7906122
U_{pp} (eV)	-57.1675810	-62.2984508	5.1308698
β_s (eV)	-20.2991100	-16.4664572	-3.8326528
β_p (eV)	-18.2386660	-21.5738169	3.3351509
ζ_s (eV)	2.3154100	2.3154100	0.0000000
ζ_p (eV)	2.1579400	2.1579400	0.0000000
α (Å$^-$1)	2.9472860	3.5367432	-0.5894572
G_{ss} (eV)	13.5900000	12.6784647	0.9115353
G_{sp} (eV)	12.6600000	12.0158734	0.6441266
G_{pp} (eV)	12.9800000	13.4771467	-0.4971467
$G_{pp'}$ (eV)	11.5900000	12.5467074	-0.9567074
H_{sp} (eV)	3.1400000	3.2468168	-0.1068168
K_1 (eV)	0.0252510	0.0262784	-0.0010274
L_1 (1/Å^2)	5.0000000	5.0512371	-0.0512371

Continued on next page

Table C.1 – *Continued from previous page*

Parameter (Units)	AM1	AM1-SPR	Δ
M_1 (Å)	1.5000000	1.3934498	0.1065502
K_2 (eV)	0.0289530	0.0217147	0.0072383
L_2 (1/Å^2)	5.0000000	5.0319955	-0.0319955
M_2 (Å)	2.1000000	1.9177053	0.1822947
K_3 (eV)	-0.0058060	-0.0072575	0.0014515
L_3 (1/Å^2)	2.0000000	1.9586210	0.0413790
M_3 (Å)	2.4000000	2.5254478	-0.1254478
PYR-Keto oxygen			
Parameter (Units)	AM1	AM1-SPR	Δ
U_{ss} (eV)	-97.8300000	-103.1104107	5.2804107
U_{pp} (eV)	-78.2623800	-85.7841018	7.5217218
β_s (eV)	-29.2727730	-29.8132437	0.5404707
β_p (eV)	-29.2727730	-34.6978950	5.4251220
ζ_s (eV)	3.1080320	3.1080320	0.0000000
ζ_p (eV)	2.5240390	2.5240390	0.0000000
α (Å$^-$1)	4.4553710	4.5774400	-0.1220690
G_{ss} (eV)	15.4200000	14.6050578	0.8149422
G_{sp} (eV)	14.4800000	14.1945053	0.2854947
G_{pp} (eV)	14.5200000	14.7352569	-0.2152569
$G_{pp'}$ (eV)	12.9800000	13.7832290	-0.8032290
H_{sp} (eV)	3.9400000	3.9843247	-0.0443247
K_1 (eV)	0.2809620	0.2107215	0.0702405
L_1 (1/Å^2)	5.0000000	5.0466690	-0.0466690
M_1 (Å)	0.8479180	0.8772223	-0.0293043
K_2 (eV)	0.0814300	0.0610725	0.0203575
L_2 (1/Å^2)	7.0000000	6.8445047	0.1554953
M_2 (Å)	1.4450710	1.5552077	-0.1101367

l	$\Theta_{\mathrm{HP},l}$	k_l
1	-2.08	70.0
2	-1.94	51.3
3	-1.74	42.3
4	-1.46	29.3
5	-1.21	41.4
6	-0.89	32.6
7	-0.64	36.7
8	-0.45	43.2
9	-0.19	41.2
10	-0.24	77.6
11	-0.15	89.7
12	-0.06	101.6
13	-0.01	163.5
14	0.03	113.8
15	0.01	63.0
16	0.16	59.1
17	0.26	47.5
18	0.47	47.3
19	0.50	23.5
20	1.03	45.4
21	1.24	34.3
22	1.54	37.3
23	1.71	63.7
24	1.88	59.3
25	2.03	73.4

Table C.2: The umbrella sampling parameters $\{\Theta_{\mathrm{HP},l}\}$ and $\{k_l\}$ in units of Å and kcal mol^{-1} Å^{-2}, respectively.